

Please type a plus sign (+) inside this box → ☐

PTO/SB/05 (12/97) (modified)
Approved for use through 09/30/00. OMB 0651-0032
Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE

**UTILITY
PATENT APPLICATION
TRANSMITTAL**

Attorney Docket No.

381552000200

Total Pages

99

First Named Inventor or Application Identifier

William J. COOK et al.

VIA HAND DELIVERY

CERTIFICATE OF HAND DELIVERY

I hereby certify that this correspondence is being hand filed with the United States Patent and Trademark Office in Washington, D.C. on February 29, 2000.

Sherri N. Shipe
Sherri N. Shipe

APPLICATION ELEMENTS

See MPEP chapter 600 concerning utility patent application contents.

ADDRESS TO:

Assistant Commissioner for Patents
Box Patent Application
Washington, DC 20231

1. ☒ Fee Transmittal Form
(Submit an original, and a duplicate for fee processing)
2. ☒ Specification [Total Pages]
(preferred arrangement set forth below)
- Descriptive title of the invention
 - Cross References to Related Applications
 - Statement Regarding Fed sponsored R & D
 - Reference to Microfiche Appendix
 - Background of the invention
 - Brief Summary of the invention
 - Brief Description of the Drawings (if filed)
 - Detailed Description
 - Claim(s)
 - Abstract of the Disclosure
3. ☒ Drawing(s) (35 USC 113) [Total Sheets]
4. ☒ Oath or Declaration (unexecuted) [Total Pages]
- a. ☐ Newly executed (original or copy)
 - b. ☐ Copy from a prior application (37 CFR 1.63(d)
(for continuation/divisional with Box 17 completed)
[Note Box 5 below]
 - i. ☐ **DELETION OF INVENTOR(S)**
Signed statement attached deleting inventor(s) named in
the prior application, see 37 CFR 1.63(d)(2) and 1.33(b)
5. ☐ Incorporation By Reference (useable if Box 4b is checked)
The entire disclosure of the prior application, from which a copy of the
oath or declaration is supplied under Box 4b, is considered as being
part of the disclosure of the accompanying application and is hereby
incorporated by reference therein

6. ☐ Microfiche Computer Program (Appendix)
7. Nucleotide and/or Amino Acid Sequence Submission
(if applicable, all necessary)
- a. ☐ Computer Readable Copy
 - b. ☐ Paper Copy (identical to computer copy)
 - c. ☐ Statement verifying identity of above copies

ACCOMPANYING APPLICATION PARTS

8. ☐ Assignment Papers (cover sheet & document(s))
9. ☐ 37 CFR 3.73(b) Statement ☐ Power of Attorney
(when there is an assignee)
10. ☐ English Translation Document (if applicable)
11. ☐ Information Disclosure Statement (IDS)/PTO-1449 ☐ Copies of IDS Citations
12. ☐ Preliminary Amendment
13. ☒ Return Receipt Postcard (MPEP 503)
(Should be specifically itemized)
14. ☐ Small Entity ☐ Statement filed in prior application,
Statement(s) Status still proper and desired
15. ☐ Certified Copy of Priority Document(s)
(if foreign priority is claimed)
16. ☐

17. If a CONTINUING APPLICATION, check appropriate box and supply the requisite information:

☐ Continuation ☐ Divisional ☐ Continuation-in-part (CIP) of prior application No:

18. CORRESPONDENCE ADDRESS

Carolyn A. Favorito
Registration No. 39,183

Morrison & Foerster LLP
2000 Pennsylvania Avenue, N.W.
Washington, D.C. 20006-1888
Telephone: (202) 887-1546
Facsimile: (202) 887-0763

- ☒ If a paper is untimely filed in the above-referenced application by applicant or his/her representative, the Assistant Commissioner is hereby petitioned under 37 C.F.R. § 1.136(a) for the minimum extension of time required to make said paper timely. In the event a petition for extension of time is made under the provisions of this paragraph, the Assistant Commissioner is hereby requested to charge any fee required under 37 C.F.R. § 1.17(a)-(d) to **Deposit Account No. 03-1952**. However, the Assistant Commissioner is **NOT** authorized to charge the cost of the issue fee to the Deposit Account.

The filing fee has been calculated as follows:

FOR	NUMBER FILED	NUMBER EXTRA	RATE	CALCULATIONS
TOTAL CLAIMS	30 - 20 =	10	x \$18.00	\$0
INDEPENDENT CLAIMS	8 - 3 =	5	x \$78.00	\$0
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			+ \$260.00	\$0
			BASIC FEE	\$690.00
			TOTAL OF ABOVE CALCULATIONS =	\$0
Reduction by 1/2 for filing by small entity (Note 37 C.F.R. §§ 1.9, 1.27, 1.28). If applicable, verified statement must be attached.				\$0
Assignment Recording Fee (if enclosed)				\$0
			TOTAL =	\$0

- ☒ **Payment of the filing fee is deferred** until receipt of the Notice to File Missing Parts.

Applicant(s) hereby petitions for any required relief including extensions of time and authorizes the Assistant Commissioner to charge the cost of such petitions and/or other fees or to credit any overpayment to **Deposit Account No. 03-1952** referencing docket no. 381552000200. A duplicate copy of this transmittal is enclosed, for that purpose.

Dated: February 29, 2000

Respectfully submitted,

By: Carolyn A. Favorito
Carolyn A. Favorito
Registration No. 39,183

Morrison & Foerster LLP
2000 Pennsylvania Avenue, N.W.
Washington, D.C. 20006-1888
Telephone: (202) 887-1546
Facsimile: (202) 887-0763

14790, A NOVEL PROTEIN KINASE MOLECULE AND USES THEREFOR

Background of the Invention

5 Phosphate tightly associated with protein has been known since the late nineteenth century. Since then, a variety of covalent linkages of phosphate to proteins have been found. The most common involve esterification of phosphate to serine, threonine, and tyrosine with smaller amounts being linked to lysine, arginine, histidine, aspartic acid, glutamic acid, and cysteine. The occurrence of phosphorylated proteins indicates the
10 existence of one or more protein kinases capable of phosphorylating amino acid residues on proteins, and also of protein phosphatases capable of hydrolyzing phosphorylated amino acid residues on proteins.

 Kinases play a critical role in the mechanism of intracellular signal transduction. They act on the hydroxyamino acids of target proteins to catalyze the transfer of a high
15 energy phosphate group from adenosine triphosphate (ATP). This process is known as protein phosphorylation. Along with phosphatases, which remove phosphates from phosphorylated proteins, kinases participate in reversible protein phosphorylation. Reversible phosphorylation acts as the main strategy for regulating protein activity in eukaryotic cells.

20 Protein kinases play critical roles in the regulation of biochemical and morphological changes associated with cell proliferation, differentiation, growth and division (D'Urso, G. et al. (1990) *Science* 250: 786-791; Birchmeier, C. et al. (1993) *Bioessays* 15: 185-189). They serve as growth factor receptors and signal transducers and have been implicated in cellular transformation and malignancy (Hunter, T. et al. (1992) *Cell* 70: 375-387; Posada, J. et al. (1992) *Mol. Biol. Cell* 3: 583-592; Hunter, T. et al. (1994) *Cell* 79: 573-582). For example, protein kinases have been shown to participate in the transmission of signals from growth-factor receptors (Sturgill, T. W. et al. (1988) *Nature* 344: 715-718; Gomez, N. et al. (1991) *Nature* 353: 170-173), cell cycle progression and control of entry of cells into mitosis (Nurse, P. (1990) *Nature* 344: 503-508; Maller, J. L. (1991) *Curr. Opin. Cell Biol.* 3: 269-275) and regulation of actin bundling (Husain-
25 Chishti, A. et al. (1988) *Nature* 334: 718-721).

 Kinases vary widely in their selectivity and specificity of target proteins. They still may, however, comprise the largest known enzyme superfamily. Protein kinases can be divided into two main groups based on either amino acid sequence similarity or specificity
35 for either serine/threonine or tyrosine residues. Serine/threonine specific kinases are often referred to as STKs while tyrosine specific kinases are referred to as PTKs. A small number of dual-specificity kinases are structurally like the serine/threonine-specific group. Within the broad classification, kinases can be further sub-divided into families whose members share a higher degree of catalytic domain amino acid sequence identity and also
40 have similar biochemical properties. Most protein kinase family members also share

structural features outside the kinase domain that reflect their particular cellular roles. These include regulatory domains that control kinase activity or interaction with other proteins (Hanks, S.K. et al. (1988) *Science* 241: 42-52).

Almost all kinases contain a catalytic domain composed of 250-300 conserved amino acids. This catalytic domain may be viewed as composed of 11 subdomains. Some of these subdomains apparently contain distinct amino acid motifs which confer specificity as a STK or PTK or both. Kinases may also contain additional amino acid sequences, usually between 5 and 100 residues, flanking or occurring within the catalytic domain. These residues apparently act to regulate kinase activity and to determine substrate specificity. (Reviewed in Hardie, G. and Hanks, S. (1995) *The Protein Kinase Facts Book*, Vol I:7-20 Academic Press, San Diego, Calif.).

Approximately one third of the known oncogenes encode PTKs. PTKs may occur as either transmembrane or soluble proteins. Transmembrane PTKs act as receptors for many growth factors. Interaction of a growth factor to its cognate receptor initiates the phosphorylation of specific tyrosine residues in the receptor itself as well as in certain second messenger proteins. Growth factors found to associate with such PTK receptors include epidermal growth factor, platelet-derived growth factor, fibroblast growth factor, hepatocyte growth factor, insulin and insulin-like growth factors, nerve growth factor, vascular endothelial growth factor, and macrophage colony stimulating factor.

Soluble PTKs often interact with the cytosolic domains of plasma membrane receptors. Receptors that signal through such PTKs include cytokine, hormone, and antigen-specific lymphocytic receptors. Many PTKs were identified as oncogene products by the observation that PTK activation was no longer subject to normal cellular controls. Also, increased tyrosine phosphorylation activity is often observed in cellular transformation, or oncogenesis, (Carbonneau, H. and Tonks, N. K. (1992) *Annu. Rev. Cell Biol.* 8:463-93.) PTK regulation may therefore be an important strategy in controlling some types of cancer.

One example of regulation of a cellular function by reversible protein phosphorylation is in the case of eukaryotic initiation factor-2 (eIF-2). When phosphorylated by an eIF-2 kinase at the alpha subunit, eIF-2 is inhibited from continuing to participate in the initiation of protein translation, which leads to termination of protein synthesis. Thus inhibitors of eIF-2 phosphorylation are expected to be anti-proliferative in nature. This regulation of eIF-2 apparently plays a role in eukaryotic cells under viral infection, nutritional deprivation, and heat shock conditions. Phosphorylation of eIF-2 apparently also plays a role in programmed cell death.

Phosphorylation of eIF-2 may be controlled by regulating the eIF-2 kinase, which is activated by double-stranded RNA (dsI). Double stranded RNA is induced by interferon

and represents an interferon mediated response to viral infection. Thus inhibitors of eIF-2 phosphorylation are expected to also be anti-viral in nature.

Summary of the Invention

5 The present invention is based, at least in part, on the discovery of novel kinases referred to herein as 14790 proteins with similarities to known eIF-2 kinases, as well as nucleic acid molecules encoding the kinase. The kinase nucleic acid and protein molecules of the present invention are useful as modulating agents in regulating a variety of cellular processes, e.g., regulation of cell cycle, including cell proliferation, differentiation, growth
10 and division. In particular, the kinase and its related nucleic acids will be advantageous in the regulation of any cellular function uncontrolled proliferation and differentiation, such as in cases of cancer. Other situations where the kinases of the invention are of particular advantage are in cases of autoimmune disorders or undesired inflammation.

 Additionally, the kinases of the present invention, similar to known eIF-2 kinases,
15 may play a role in cellular metabolism in response to conditions such as viral infection, nutritional deprivation, and heat shock conditions. As such, inhibitory modulators of the kinases are expected to be of benefit as anti-viral agents. A kinase of the invention has been observed to be up-regulated in HBV infected liver cells. Thus modulators of the kinases of the invention may be used in cases of HBV infection.

20 The kinases of the invention are also expected to play a role in cardiac cellular processes, either independent of, or in connection with, programmed cell death (apoptosis).

 Thus, in one aspect, this invention provides isolated nucleic acid molecules encoding 14790 proteins or biologically active portions thereof, as well as nucleic acid fragments suitable as primers or hybridization probes for the detection of kinase-encoding
25 nucleic acids.

 In one embodiment, a kinase encoding a nucleic acid molecule of the invention is at 75%, 77%, 80%, 82%, 85%, 87%, 90%, 92%, 95%, 97%, 98%, 99% or greater homology to a nucleotide sequence (e.g., to the entire length of the nucleotide sequence) including SEQ ID NO:1 or a complement thereof. In a preferred embodiment, the isolated
30 nucleic acid molecule includes the nucleotide sequence shown SEQ ID NO:1 or a coding region of SEQ ID NO:1, or a complement thereof. In another embodiment, the nucleic acid molecule includes the 5' UTR and the coding region of SEQ ID NO:1. In yet another embodiment, the nucleic acid molecule includes the coding region of SEQ ID NO:3 and the 3' UTR of SEQ ID NO:1. In another preferred embodiment, the nucleic acid molecule
35 has the nucleotide sequence shown in SEQ ID NO:1 or the coding region of SEQ ID NO:1. In another preferred embodiment, the nucleic acid molecule comprises a fragment of at least 4400 nucleotides of the nucleotide sequence of SEQ ID NO:1 or the coding region of SEQ ID NO:1, or a complement thereof.

In another embodiment, a kinase nucleic acid molecule includes a nucleotide sequence encoding a protein having an amino acid sequence sufficiently homologous to the amino acid sequence of SEQ ID NO:2. In a preferred embodiment, a kinase nucleic acid molecule includes a nucleotide sequence encoding a protein having an amino acid sequence at least 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more homologous to an amino acid sequence including SEQ ID NO:2 (e.g., the entire amino acid sequence of SEQ ID NO:2).

In another preferred embodiment, an isolated nucleic acid molecule encodes the amino acid sequence of a human kinase, preferably with similarities to known eIF-2 kinases. In yet another preferred embodiment, the nucleic acid molecule includes a nucleotide sequence encoding a protein which includes the amino acid sequence of SEQ ID NO: 2. In yet another preferred embodiment, the nucleic acid molecule includes a nucleotide sequence encoding a protein having the amino acid sequence of SEQ ID NO: 2.

Another embodiment of the invention features nucleic acid molecules, preferably kinase nucleic acid molecules, which specifically detect kinase nucleic acid molecules relative to other nucleic acid molecules. In one embodiment, the nucleic acid molecules of the invention may be used to detect and/or differentiate between different eIF-2 kinase encoding nucleic acids. For example, in one embodiment, such a nucleic acid molecule is at least 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, or 5500 nucleotides in length and hybridizes under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence shown in SEQ ID NO:1 or a complement thereof.

In other preferred embodiments, the nucleic acid molecule encodes a naturally occurring allelic variant of a polypeptide which includes the amino acid sequence of SEQ ID NO:2, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule which includes SEQ ID NO:1 or the coding region of SEQ ID NO:1 under stringent conditions.

Another embodiment of the invention provides an isolated nucleic acid molecule which is antisense to the nucleic acid molecule of the invention, e.g., the coding strand of a nucleic acid molecule of the invention.

Another aspect of the invention provides a vector comprising the nucleic acid molecule. In certain embodiments, the vector is a recombinant expression vector. In another embodiment, the invention provides a host cell containing a vector of the invention. The invention also provides a method for producing a kinase, preferably a protein, by culturing in a suitable medium, a host cell, e.g., a mammalian host cell such as a non-human mammalian cell, of the invention containing a recombinant expression vector, such that the protein is produced.

Another aspect of this invention features isolated or recombinant proteins and polypeptides. In one preferred embodiment, the isolated protein, preferably a 14790 protein

includes at least one N-glycosylation site; at least one cGMP-dependent protein kinase phosphorylation site; at least one protein kinase C phosphorylation site; at least one casein kinase II phosphorylation site; at least one tyrosine kinase phosphorylation site; at least one N-myristoylation site; at least one amidation site; at least one protein kinase ATP-binding region signature; and at least one Ser/Thr protein kinase active-site signature; and at least one DNA polymerase family B signature.

In another embodiment, the isolated protein, preferably a 14790 protein, includes an amino acid sequence which is of at least 51%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 85%, 90%, 95%, 99% or greater homology to an amino acid sequence including SEQ ID NO:2. Preferably, the proteins are kinases.

In yet another embodiment, the isolated protein, preferably a 14790 protein, is expressed and/or functions in cells of the hepatic system. Preferably, such proteins act as kinases.

In an even further embodiment, the isolated protein, preferably a 14790 protein, plays a role in signalling pathways associated with cellular growth, e.g., signalling pathways associated with cell cycle regulation and act as kinases.

In another embodiment, the isolated protein, preferably a 14790 protein, includes at least one N-glycosylation site; at least one cGMP-dependent protein kinase phosphorylation site; at least one protein kinase C phosphorylation site; at least one casein kinase II phosphorylation site; at least one tyrosine kinase phosphorylation site; at least one N-myristoylation site; at least one amidation site; at least one protein kinase ATP-binding region signature; and at least one Ser/Thr protein kinase active-site signature; and at least one DNA polymerase family B signature and is able to phosphorylate the core protein of Hepatitis B Virus (HBV).

In another embodiment, the isolated protein, preferably a 14790 protein, is encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1. Preferably, the proteins are kinases.

In another embodiment, the isolated protein, preferably a 14790 protein, has an amino acid sequence homologous to the amino acid sequence of SEQ ID NO:2. In a preferred embodiment, the protein, preferably a 14790 protein, has an amino acid sequence at least about 50%, 55%, 59%, 60%, 65%, 70%, 75%, 80%, 81%, 85%, 90%, 95%, 98% or greater homology to an amino acid sequence including SEQ ID NO:2 (e.g., the entire amino acid sequence of SEQ ID NO:2). In another embodiment, the invention features fragments of the proteins having the amino acid sequence of SEQ ID NO:2, wherein the fragment comprises at least 15, 17, 19 or 21 amino acids (e.g., contiguous amino acids) of the amino acid sequence of SEQ ID NO:2. In another embodiment, the protein, preferably a 14790 protein, has the amino acid sequence of SEQ ID NO:2.

Another embodiment of the invention features an isolated protein, preferably a 14790 protein, which is encoded by a nucleic acid molecule having a nucleotide sequence at least about 50%, 54%, 55%, 60%, 62%, 65%, 70%, 75%, 78%, 80%, 85%, 86%, 90%, 95%, 97%, 98% or greater homology to a nucleotide sequence (e.g., to the entire length of the nucleotide sequence) including SEQ ID NO:1 or the coding region of SEQ ID NO:1 a complement thereof. This invention further features an isolated protein, preferably a 14790 protein, which is encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or a complement thereof.

The proteins of the present invention or biologically active portions thereof, can be operatively linked to a non-14790 polypeptide (e.g., heterologous amino acid sequences) to form fusion proteins. The invention further features antibodies, such as monoclonal or polyclonal antibodies, that specifically bind proteins of the invention, preferably 14790 proteins. In addition, the 14790 proteins or biologically active portions thereof can be incorporated into pharmaceutical compositions, which optionally include pharmaceutically acceptable carriers.

In another aspect, the present invention provides a method for detecting the presence of a kinase nucleic acid molecule, protein or polypeptide in a biological sample by contacting the biological sample with one or more agent(s) capable of detecting a kinase nucleic acid molecule, protein or polypeptide such that the presence of a kinase nucleic acid molecule, protein or polypeptide is detected in the biological sample. Examples of agents for the detection of kinases and nucleic acids that encode them are well known in the art.

In another aspect, the present invention provides a method for detecting the presence of kinase activity in a sample by contacting the biological sample with one or more agent(s) capable of detecting 14790 activity. Such agents are known in the art. Examples of agents useful for the detection of eIF-2 kinase activity or the like include peptides or proteins containing eIF-2 phosphorylation target sequences.

In another aspect, the invention provides a method for modulating 14790 activity in a cell by contacting a cell capable of expressing kinase with an agent that modulates kinase activity. In one embodiment, the modulating agent inhibits kinase activity. In another embodiment, the modulating agent stimulates or increases kinase activity. In one embodiment of an inhibitory modulator, the agent is an antibody that specifically binds to a 14790 protein. In another embodiment, the modulating agent regulates expression of kinase by modulating transcription of a kinase gene or translation of a kinase mRNA. In yet another embodiment, the agent is a nucleic acid molecule having a nucleotide sequence that is antisense to the coding strand of a kinase mRNA or a kinase gene.

In one embodiment, the methods of the present invention are used to treat a subject having a disorder characterized by aberrant 14790 protein activity or nucleic acid expression by administering an agent which is a kinase modulator to the subject. In one embodiment, the kinase modulator is a protein. In another embodiment the kinase related modulator is a kinase nucleic acid molecule. In yet another embodiment, the kinase modulator is a peptide, peptidomimetic, or other small molecule. In a preferred embodiment, the disorder characterized by aberrant 14790 protein or nucleic acid expression is a disorder characterized by enhanced viral replication, e.g. during infection with HBV.

The present invention also provides a diagnostic assay for identifying the presence or absence of a genetic alteration characterized by at least one of (i) aberrant modification or mutation of a gene encoding a 14790 protein; (ii) mis-regulation of the gene; and (iii) aberrant post-translational modification of a 14790 protein, wherein a wild-type form of the gene encodes a protein with a kinase activity.

In another aspect the invention provides a method for identifying a compound that binds to or modulates the activity of a 14790 protein, by providing an indicator composition comprising a 14790 protein having kinase activity, contacting the indicator composition with a test compound, and determining the effect of the test compound on kinase activity in the indicator composition to identify a compound that modulates the activity of a 14790 protein.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

Brief Description of the Drawings

Figure 1 depicts the cDNA sequence and predicted amino acid sequence of human kinase. The nucleotide sequence corresponds to the 5525 nucleic acids of SEQ ID NO:1 which include nucleic acids 1-4950 of the coding region, the 5' UTR of 62 nucleic acids, and the 3' UTR of 513 nucleic acids. The amino acid sequence corresponds to amino acids 1 to 1650 of SEQ ID NO:2.

Figures 2 a-c shows a multiple sequence alignment of the amino acid sequence of SEQ ID NO:2 in comparison with known mouse and human kinase

Figure 3 is a graph of TaqMan RT-PCR data illustrating the relative gene expression on HBV positive liver using liver PTI as a reference sample.

Figure 4 is a graph of TaqMan RT-PCR data illustrating the relative gene expression on HepG2.2.15 (HBV-expressing line) using liver PTI as a reference sample.

Figure 5 is a graph of data illustrating the relative gene expression on in normal human tissue using thyroid as a reference

Figure 6 depicts the structural components of the protein of the amino acid sequence as generated from Protean software.

Figures 7 a-n comprise data generated to show PFAM cites, hydrophobicity/hydrophilicity, and cysteine residues of the amino acid sequence of SEQ ID NO:2, as well as PSORT prediction of protein localization, signal peptide predictions, transmembrane segments predicted by MEMSAT, Prosite pattern matches, protein family/domain matches and ProDom matches of the amino acid sequence of SEQ ID NO:2.

10 **Detailed Description of the Invention**

The present invention is based, at least in part, on the discovery of a novel 14790 molecule, referred to herein as "kinase" or "kinase" nucleic acid and a polypeptide molecule, which play a role in or function in signalling pathways associated with cellular growth. In one embodiment, the molecule modulates the activity of one or more proteins involved in cellular growth or differentiation, e.g., hepatic cell growth or differentiation. In another embodiment, the molecule of the present invention has its mRNA induced in liver cells infected with HBV. In another embodiment, the molecule of the present invention encodes a polypeptide that is able to phosphorylate the core protein of HBV.

In a preferred embodiment, the molecules are protein kinases which are expressed and/or function in cells of the hepatic system, e.g., cells of the liver and the associated blood vessels of the liver.

As used herein, the term "protein kinase" includes a protein or polypeptide which is capable of modulating its own phosphorylation state or the phosphorylation state of another protein or polypeptide. Protein kinases can have a specificity for (i.e., a specificity to phosphorylate) serine/threonine residues, tyrosine residues, or both serine/threonine and tyrosine residues, e.g., the dual specificity kinases. As referred to herein, protein kinases may include a catalytic domain of about 150-400 amino acid residues in length, preferably about 170-300 amino acid residues in length, or more preferably about 190-300 amino acid residues in length, which includes preferably 5-20, more preferably 5-15, or preferably 11 highly conserved motifs or subdomains separated by sequences of amino acids with reduced or minimal conservation. Specificity of a protein kinase for phosphorylation of either tyrosine or serine/threonine can be predicted by the sequence of two of the subdomains (VIb and VIII) in which different residues are conserved in each class (as described in, for example, Hanks et al. (1988) Science 241:42-52) the contents of which are incorporated herein by reference). These subdomains are also described in further detail herein.

Protein kinases play a role in signalling pathways associated with cellular growth. For example, protein kinases are involved in the regulation of signal transmission from

cellular receptors, e.g., growth-factor receptors; entry of cells into mitosis; and the regulation of cytoskeleton function, e.g., actin bundling. Thus, the molecules of the present invention may be involved in: 1) the regulation of transmission of signals from cellular receptors, e.g., cardiac cell growth factor receptors; 2) the modulation of the entry of cells, e.g., cardiac precursor cells, into mitosis; 3) the modulation of cellular differentiation; 4) the modulation of cell death; and 5) the regulation of cytoskeleton function, e.g., actin bundling.

Inhibition or over stimulation of the activity of protein kinases involved in signaling pathways associated with cellular growth can lead to perturbed cellular growth, which can in turn lead to cellular growth related disorders. As used herein, a "cellular growth related disorder" includes a disorder, disease, or condition characterized by a deregulation, e.g., an upregulation or a downregulation, of cellular growth. Cellular growth deregulation may be due to a deregulation of cellular proliferation, cell cycle progression, cellular differentiation and/or cellular hypertrophy. Examples of cellular growth related disorders include cardiovascular disorders such as heart failure, hypertension, atrial fibrillation, dilated cardiomyopathy, idiopathic cardiomyopathy, or angina; proliferative disorders or differentiative disorders such as cancer, e.g., melanoma, prostate cancer, cervical cancer, breast cancer, colon cancer, or sarcoma or fibrotic lesions seen in liver fibrosis.

The present invention is based, at least in part, on the discovery of novel molecules, referred to herein as protein and nucleic acid molecules, which comprise a family of molecules having certain conserved structural and functional features. The term "family" when referring to the protein and nucleic acid molecules of the invention is intended to mean two or more proteins or nucleic acid molecules having a common structural domain or motif and having sufficient amino acid or nucleotide sequence homology as defined herein. Such family members can be naturally or non-naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of human origin, as well as other, distinct proteins of human origin or alternatively, can contain homologues of non-human origin. Members of a family may also have common functional characteristics.

One embodiment of the invention features a nucleic acid molecule, preferably a human 14790 molecule, which was identified from a cDNA library. The nucleic acid and protein molecule of the invention is described in further detail in the following subsections.

A. The Kinase Nucleic Acid and Protein Molecules

In one embodiment, the isolated proteins of the present invention, preferably 14790 proteins, are identified based on the presence of at least one "Ser/Thr kinase site" and at least one "ATP-binding region." As used herein, the term "Ser/Thr kinase site" includes an amino acid sequence of about 200-400 amino acid residues in length, preferably 200-300 amino acid residues in length, and more preferably 250-300 amino acid residues in length, which is conserved in kinases which phosphorylate serine and threonine residues and found in the catalytic domain of Ser/Thr kinases. Preferably, the Ser/Thr kinase site includes the following amino acid consensus sequence X₉-g-X-G-X₄-V-X₁₂-K-X-(10-19)-E-X₆₆-h-X₈-h-r-D-X-K-X₂-N-X₁₇-K-X₂-D-f-g-X₂₁-p-X₁₃-w-X₃-g-X₅₅-R-X₁₄-h-X₃ (where invariant residues are indicated by upper case letters and nearly invariant residues are indicated by lower case letters). The nearly invariant residues are usually found in most Ser/Thr kinase sites, but can be replaced by other amino acids which, preferably, have similar characteristics. For example, a nearly invariant hydrophobic amino acid in the above amino acid consensus sequence would most likely be replaced by another hydrophobic amino acid. Ser/Thr kinase domains are described in, for example, Levin D.E. et al. (1990) Proc. Natl. Acad. Sci. USA 87:8272-76, the contents of which are incorporated herein by reference.

As used herein, the term "ATP-binding region" includes an amino acid sequence of about 5-40, preferably 5-25, and more preferably 5-15 amino acid residues in length, present in enzymes which activate their substrates by phosphorylation, and involved in binding adenosine triphosphate (ATP). ATP-binding regions preferably include the following amino acid consensus sequence: G-X-G-X-X-G-X(15-23)-K. ATP-binding regions are described in, for example, Samuel K.P. et al. (1987) FEBS Let. 218(1): 81-86, the contents of which are incorporated herein by reference. Amino acid residues 596-604 of kinase comprise an ATP-binding region.

Isolated proteins of the present invention, preferably 14790 proteins, have an amino acid sequence sufficiently homologous to the amino acid sequence of SEQ ID NO:2 or are encoded by a nucleotide sequence sufficiently homologous to SEQ ID NO:1. As used herein, the term "sufficiently homologous" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (e.g., an amino acid residue which has a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences share common structural domains or motifs and/or a common functional activity. For example, amino acid or nucleotide sequences which share common structural domains have at least 30%, 40%, or 50% homology, preferably 60% homology, more preferably 70%-80%, and even more preferably 90-95% homology across the amino acid sequences of the domains and contain at least one and preferably two structural domains or motifs, are defined herein as

sufficiently homologous. Furthermore, amino acid or nucleotide sequences which share at least 30%, 40%, or 50%, preferably 60%, more preferably 70-80%, or 90-95% homology and share a common functional activity are defined herein as sufficiently homologous.

As used interchangeably herein a "kinase activity", "biological activity of kinase" or "functional activity of kinase", refers to an activity exerted by a 14790 protein, polypeptide or nucleic acid molecule on a kinase responsive cell or a 14790 protein substrate, as determined in vivo, or in vitro, according to standard techniques. The biological activity of kinase is described herein.

Accordingly, another embodiment of the invention features isolated 14790 proteins and polypeptides having a kinase activity. Preferred proteins are 14790 proteins having at least one N-glycosylation site; at least one cGMP-dependent protein kinase phosphorylation site; at least one protein kinase C phosphorylation site; at least one casein kinase II phosphorylation site; at least one tyrosine kinase phosphorylation site; at least one N-myristoylation site; at least one amidation site; at least one protein kinase ATP-binding region signature; and at least one Ser/Thr protein kinase active-site signature; and at least one DNA polymerase family B signature and, preferably, a kinase activity. Additional preferred proteins have at least one N-glycosylation site; at least one cGMP-dependent protein kinase phosphorylation site; at least one protein kinase C phosphorylation site; at least one casein kinase II phosphorylation site; at least one tyrosine kinase phosphorylation site; at least one N-myristoylation site; at least one amidation site; at least one protein kinase ATP-binding region signature; and at least one Ser/Thr protein kinase active-site signature; and at least one DNA polymerase family B signature and are, preferably, encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or the coding region thereof.

The nucleotide sequence of the isolated human kinase cDNA and the predicted amino acid sequence of the human 14790 polypeptide are shown in Figure 1 and in SEQ ID NOs:1 and 2, respectively. A plasmid containing the nucleotide sequence encoding human kinase was deposited with American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, VA 20110-2209, on _____ and assigned Accession Number _____. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

The kinase gene, which is approximately 5525 nucleotides in length, encodes a protein having a molecular weight of approximately 181.5 kD and which is approximately 1650 amino acid residues in length. The kinase gene is expressed predominantly in skeletal muscle, brain and liver.

Various aspects of the invention are described in further detail in the following subsections:

5 I. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode 14790 proteins or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes to identify kinase-encoding nucleic acids (e.g., kinase mRNA) and fragments for use as PCR primers for the amplification or mutation of 10 kinase nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

15 An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term "isolated" includes nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic 20 acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated kinase nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. 25 Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule 30 having the nucleotide sequence of SEQ ID NO:1 or coding region thereof, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. For example, using all or portion of the nucleic acid sequence of SEQ ID NO:1, or the coding region thereof, as a hybridization probe, kinase nucleic acid molecules can be isolated using standard hybridization and cloning techniques 35 (e.g., as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Moreover, a nucleic acid molecule encompassing all or a portion of SEQ ID NO:1, or the coding region thereof, can be isolated by the polymerase chain reaction (PCR) using synthetic oligonucleotide primers designed based upon the sequence of SEQ ID NO:1, or the coding region thereof.

5 A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to nucleotide sequences can be prepared by
10 standard synthetic techniques, e.g., using an automated DNA synthesizer.

In a preferred embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO:1. The sequence of SEQ ID NO:1 corresponds to the partial human kinase cDNA. This cDNA comprises sequences encoding the human 14790 protein (i.e., "the coding region", from nucleotides 1-4950), as
15 well as 5' untranslated sequences (62 nucleotides) and 3' untranslated sequences (513 nucleotides). Alternatively, the nucleic acid molecule can comprise only the coding region of SEQ ID NO:1 (e.g., nucleotides 1-4950).

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence
20 shown in SEQ ID NO:1, or the coding region thereof, or a portion of any of these nucleotide sequences. A nucleic acid molecule which is complementary to the nucleotide sequence shown in SEQ ID NO:1, or the coding region thereof, is one which is sufficiently complementary to the nucleotide sequence shown in SEQ ID NO:1, or the coding region thereof, such that it can hybridize to the nucleotide sequence shown in SEQ ID NO:1, or
25 the coding region thereof, thereby forming a stable duplex.

In still another preferred embodiment, an isolated nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least about 50%, 54%, 55%, 60%, 62%, 65%, 70%, 75%, 78%, 80%, 85%, 86%, 90%, 95%, 97%, 98% or more homologous to the nucleotide sequence (e.g., to the entire length of the nucleotide
30 sequence) shown in SEQ ID NO:1, or the coding region thereof, or a portion of any of these nucleotide sequences.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of SEQ ID NO:1, or the coding region thereof, for example a fragment which can be used as a probe or primer or a fragment encoding a biologically
35 active portion of a 14790 protein. The nucleotide sequence determined from the cloning of the kinase gene allows for the generation of probes and primers designed for use in identifying and/or cloning other kinase family members, as well as kinase homologues from other species. The probe/primer typically comprises substantially purified

oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12 or 15, preferably about 20 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, or 75 consecutive nucleotides of a sense sequence of SEQ ID NO:1, of an anti-sense sequence of SEQ ID NO:1, or the coding region thereof, or of a naturally occurring allelic variant or mutant of SEQ ID NO:1, or the coding region thereof. In an exemplary embodiment, a nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1500, 2000, 2500, 3000, 3500, 4000, or 4500 nucleotides in length and hybridizes under stringent hybridization conditions to a nucleic acid molecule of SEQ ID NO:1, or the coding region thereof.

Probes based on the nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In preferred embodiments, the probe further comprises a label group attached thereto, e.g., the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissues which misexpress a 14790 protein, such as by measuring a level of a kinase-encoding nucleic acid in a sample of cells from a subject e.g., detecting kinase mRNA levels or determining whether a genomic kinase gene has been mutated or deleted.

A nucleic acid fragment encoding a "biologically active portion of a 14790 protein" can be prepared by isolating a portion of the nucleotide sequence of SEQ ID NO:1, or the coding region thereof, which encodes a polypeptide having a kinase biological activity (the biological activities of the 14790 proteins are described herein), expressing the encoded portion of the 14790 protein (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the 14790 protein.

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence shown in SEQ ID NO:1, or the coding region thereof, due to the degeneracy of the genetic code and, thus, encode the same 14790 proteins as those encoded by the nucleotide sequence shown in SEQ ID NO:1, or the coding region thereof. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NO:2.

In addition to the kinase nucleotide sequences shown in SEQ ID NO:1, or the coding region thereof, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the 14790 proteins may exist within a population (e.g., the human population). Such genetic polymorphism in the kinase genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules which include an open reading frame encoding a 14790 protein, preferably a mammalian 14790 protein, and can further include non-coding regulatory sequences, and

introns. Such natural allelic variations include both functional and non-functional proteins and can typically result in 1-5% variance in the nucleotide sequence of a kinase gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in genes that are the result of natural allelic variation and that do not alter the functional activity of a protein are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding other kinase family members and, thus, which have a nucleotide sequence which differs from the kinase sequences of SEQ ID NO:1, or the coding region thereof, are intended to be within the scope of the invention. For example, another kinase cDNA can be identified based on the nucleotide sequence of human kinase. Moreover, nucleic acid molecules encoding proteins from different species, and thus which have a nucleotide sequence which differs from the kinase sequences of SEQ ID NO:1, or the coding region thereof, are intended to be within the scope of the invention. For example, a mouse kinase cDNA can be identified based on the nucleotide sequence of a human kinase .

Nucleic acid molecules corresponding to natural allelic variants and homologues of the kinase cDNAs of the invention can be isolated based on their homology to the kinase nucleic acids disclosed herein using the cDNAs disclosed herein, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15, 20, 25, 30 or more nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, or the coding region thereof. In other embodiment, the nucleic acid is at least 30, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1500, 2000, 2500, 3000, 3500, 4000, or 4500 nucleotides in length. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 30%, 40%, 50%, or 60% homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 70%, more preferably at least about 80%, even more preferably at least about 85% or 90% homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. A more preferred example of stringent hybridization conditions is hybridization in 0.5M sodium phosphate, 7% SDS at 65°C , followed by one or more washes in 0.2 X SSC at 65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under

stringent conditions to the sequence of SEQ ID NO:1, or the coding region thereof, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

5 In addition to naturally-occurring allelic variants of the kinase sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences of SEQ ID NO:1, or the coding region thereof, thereby leading to changes in the amino acid sequence of the encoded 14790 proteins, without altering the functional ability of the 14790 proteins. For example, 10 nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NO:1, or the coding region thereof. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of kinase (e.g., the sequence of SEQ ID NO:2) without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For 15 example, amino acid residues that are conserved among the 14790 proteins of the present invention, are predicted to be particularly unamenable to alteration. Furthermore, additional amino acid residues that are conserved between the 14790 proteins of the present invention and other kinase family members are not likely to be amenable to alteration.

20 Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding 14790 proteins that contain changes in amino acid residues that are not essential for activity. Such 14790 proteins differ in amino acid sequence from SEQ ID NO:2, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an 25 amino acid sequence at least about 41%, 42%, 45%, 50%, 55%, 59%, 60%, 65%, 70%, 75%, 80%, 81%, 85%, 90%, 95%, 98% or more homologous to the amino acid sequence of SEQ ID NO:2 (e.g., the entire amino acid sequence of SEQ ID NO:2).

An isolated nucleic acid molecule encoding a 14790 protein homologous to the protein of SEQ ID NO:2 can be created by introducing one or more nucleotide 30 substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:1 such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into SEQ ID NO:1, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A 35 "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g.,

aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a 14790 protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a kinase coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for kinase biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:1, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

In a preferred embodiment, a mutant 14790 protein can be assayed for the ability to: 1) regulate transmission of signals from cellular receptors, e.g., cardiac cell growth factor receptors; 2) control entry of cells, e.g., cardiac cells, into mitosis; 3) modulate cellular differentiation; 4) modulate cell death; 5) regulate cytoskeleton function, e.g., actin bundling; or being able to phosphorylate the core protein of HBV.

In addition to the nucleic acid molecules encoding 14790 proteins described above, another aspect of the invention pertains to isolated nucleic acid molecules which are antisense thereto. An "antisense" nucleic acid comprises a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire kinase coding strand, or only to a portion thereof. In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding kinase. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues (e.g., the coding region of human kinase corresponds to 1-4950 of SEQ ID NO:1. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding kinase. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (i.e., also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding kinase disclosed herein (e.g., nucleic acids 1-4950 of SEQ ID NO:1 antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of kinase mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of kinase mRNA. For example, the antisense oligonucleotide can be

complementary to the region surrounding the translation start site of kinase mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50] nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art.

- 5 For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples
- 10 of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine,
- 15 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-
- 20 oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described
- 25 further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a 14790 protein to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. The hybridization can be by

30 conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention include direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target

35 selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or

antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

5 In yet another embodiment, the antisense nucleic acid molecule of the invention is an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the strands run parallel to each other (Gaultier et al. (1987) *Nucleic Acids. Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide
10 (Inoue et al. (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al. (1987) *FEBS Lett.* 215:327-330).

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as a kinase mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave kinase mRNA transcripts to thereby inhibit translation of kinase mRNA. A ribozyme having specificity for a kinase-encoding nucleic acid can be designed based upon the nucleotide sequence of a kinase cDNA disclosed herein (i.e., SEQ ID NO:1, or the coding region thereof). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a kinase-encoding mRNA. See, e.g., Cech et al. U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent No. 5,116,742. Alternatively, kinase mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel, D. and Szostak, J.W. (1993) *Science* 261:1411-1418.

Alternatively, kinase gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the kinase (e.g., the kinase promoter and/or enhancers) to form triple helical structures that prevent transcription of the kinase gene in target cells. See generally, Helene, C. (1991) *Anticancer Drug Des.* 6(6):569-84; Helene, C. et al. (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher, L.J. (1992) *Bioassays* 14(12):807-15.

In yet another embodiment, the kinase nucleic acid molecules of the present invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup B. et al. (1996) *Bioorganic & Medicinal Chemistry* 4 (1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid

mimics, e.g., DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup B. et al. (1996) *supra*; Perry-O'Keefe et al. Proc. Natl. Acad. Sci. 93: 14670-675.

PNAs of kinase nucleic acid molecules can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNAs of kinase nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (e.g., by PNA-directed PCR clamping); as 'artificial restriction enzymes' when used in combination with other enzymes, (e.g., S1 nucleases (Hyrup B. (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup B. et al. (1996) *supra*; Perry-O'Keefe *supra*).

In another embodiment, PNAs of kinase can be modified, (e.g., to enhance their stability or cellular uptake), by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of kinase nucleic acid molecules can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, (e.g., RNase H and DNA polymerases), to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup B. (1996) *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup B. (1996) *supra* and Finn P.J. et al. (1996) *Nucleic Acids Res.* 24 (17): 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs, e.g., 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used as a between the PNA and the 5' end of DNA (Mag, M. et al. (1989) *Nucleic Acid Res.* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn P.J. et al. (1996) *supra*). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser, K.H. et al. (1975) *Bioorganic Med. Chem. Lett.* 5: 1119-11124).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger et al. (1989) *Proc. Natl. Acad. Sci. US.* 86:6553-6556; Lemaitre et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT

Publication No. W088/09810) or the blood-brain barrier (see, e.g., PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (See, e.g., Krol et al. (1988) *Bio-Techniques* 6:958-976) or intercalating agents. (See, e.g., Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide
 5 may be conjugated to another molecule, (e.g., a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

II. Isolated 14790 proteins and Anti-Kinase Antibodies

One aspect of the invention pertains to isolated 14790 proteins, and biologically
 10 active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise anti-kinase antibodies. In one embodiment, native 14790 proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, 14790 proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a 14790 protein or
 15 polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the 14790 protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially
 20 free of cellular material" includes preparations of 14790 protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of 14790 protein having less than about 30% (by dry weight) of non-14790 protein (also referred to herein as a "contaminating protein"), more preferably less
 25 than about 20% of non-14790 protein, still more preferably less than about 10% of non-14790 protein, and most preferably less than about 5% non-14790 protein. When the 14790 protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5%
 30 of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of 14790 protein in which the protein is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. In one
 35 embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of 14790 protein having less than about 30% (by dry weight) of chemical precursors or non-kinase chemicals, more preferably less than about 20% chemical precursors or non-kinase chemicals, still more preferably less than about 10%

chemical precursors or non-kinase chemicals, and most preferably less than about 5% chemical precursors or non-kinase chemicals.

Biologically active portions of a 14790 protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the 14790 protein, e.g., the amino acid sequence shown in SEQ ID NO:2, which include less amino acids than the full length 14790 proteins, and exhibit at least one activity of a 14790 protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the 14790 protein. A biologically active portion of a 14790 protein can be a polypeptide which is, for example, at least 10, 25, 50, 100 or more amino acids in length.

In a preferred embodiment, the 14790 protein has an amino acid sequence shown in SEQ ID NO:2. In other embodiments, the 14790 protein is substantially homologous to SEQ ID NO:2, and retains the functional activity of the protein of SEQ ID NO:2, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail in subsection I above. Accordingly, in another embodiment, the 14790 protein is a protein which comprises an amino acid sequence at least about 55%, 59%, 60%, 65%, 70%, 75%, 80%, 81%, 85%, 90%, 95%, 98% or more homologous to the amino acid sequence of SEQ ID NO:2 (e.g., the entire amino acid sequence of SEQ ID NO:2).

To determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, or 90% of the length of the reference sequence (e.g., when aligning a second sequence to the kinase amino acid sequence of SEQ ID NO:2 having 1650 amino acid residues, at least 49, preferably at least 660, more preferably at least 825, even more preferably at least 990, and even more preferably at least 1155, 1320 or 1486 amino acid residues are aligned). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred

embodiment, the percent identity between two amino acid sequences is determined using the GAP program in the GCG software package (available at <http://www.gcg.com>), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at <http://www.gcg.com>), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A more preferred embodiment, the percent identity between two nucleotide or amino acid sequences is determined using a Blosum 62 scoring matrix with a gap open penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See <http://www.ncbi.nlm.nih.gov>.

The invention also provides kinase chimeric or fusion proteins. As used herein, a kinase "chimeric protein" or "fusion protein" comprises a 14790 polypeptide operatively linked to a non-14790 polypeptide. A kinase "polypeptide" refers to a polypeptide having an amino acid sequence corresponding to kinase, whereas a "non-14790 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the 14790 protein, e.g., a protein which is different from the 14790 protein and which is derived from the same or a different organism. Within a kinase fusion protein the 14790 polypeptide can correspond to all or a portion of a protein. In a preferred embodiment, a kinase fusion protein comprises at least one biologically active portion of a 14790 protein. In another preferred embodiment, a kinase fusion protein comprises at least two biologically active portions of a 14790 protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the 14790 polypeptide and the non-14790 polypeptide are fused in-frame to each other. The non-14790 polypeptide can be fused to the N-terminus or C-terminus of the 14790 polypeptide.

For example, in one embodiment, the fusion protein is a GST-kinase fusion protein in which the kinase sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant kinase.

In another embodiment, the fusion protein is a 14790 protein containing a
5 heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells), expression and/or secretion of kinase can be increased through use of a heterologous signal sequence.

The kinase fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject in vivo. The kinase fusion proteins can be
10 used to affect the bioavailability of a kinase substrate. Use of kinase fusion proteins may be useful therapeutically for the treatment of viral infections, e.g., infection of liver cells with HBV. Moreover, the kinase-fusion proteins of the invention can be used as immunogens to produce anti-kinase antibodies in a subject, to purify kinase ligands and in screening assays to identify 14790 molecules which inhibit the interaction of with a kinase
15 substrate.

Preferably, a kinase chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with
20 conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR
25 amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, Current Protocols in Molecular Biology, eds. Ausubel et al. John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A kinase-encoding nucleic acid can be cloned
30 into such an expression vector such that the fusion moiety is linked in-frame to the 14790 protein.

The present invention also pertains to variants of the 14790 proteins which function as either kinase agonists (mimetics) or as kinase antagonists. Variants of the 14790
35 proteins can be generated by mutagenesis, e.g., discrete point mutation or truncation of a 14790 protein. An agonist of the 14790 proteins can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of a 14790 protein. An antagonist of a 14790 protein can inhibit one or more of the activities of the naturally occurring form of the 14790 kinase protein. An antagonist of the 14790 protein can inhibit

one or more of the activities of the naturally occurring form of the kinase protein by, for example, by competitively modulating the ability of 14790 protein to phosphorylate the core protein of HBV.. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant
5 having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the 14790 protein.

In one embodiment, variants of a 14790 protein which function as either kinase agonists (mimetics) or as kinase antagonists can be identified by screening combinatorial
10 libraries of mutants, e.g., truncation mutants, of a 14790 protein for 14790 protein agonist or antagonist activity. In one embodiment, a variegated library of kinase variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of kinase variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene
15 sequences such that a degenerate set of potential kinase sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display) containing the set of kinase sequences therein. There are a variety of methods which can be used to produce libraries of potential kinase variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be
20 performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the desired set of potential kinase sequences. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang, S.A. (1983) Tetrahedron 39:3; Itakura et al. (1984) Annu. Rev. Biochem. 53:323; Itakura et al. (1984) Science 198:1056; Ike et al. (1983) Nucleic Acid Res. 11:477.
25

In addition, libraries of fragments of a 14790 protein coding sequence can be used to generate a variegated population of kinase fragments for screening and subsequent selection of variants of a 14790 protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a kinase coding
30 sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an
35 expression library can be derived which encodes N-terminal, C-terminal and internal fragments of various sizes of the 14790 protein.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA

libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of 14790 proteins. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into
5 replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the
10 screening assays to identify kinase variants (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave et al. (1993) *Protein Engineering* 6(3):327-331).

In one embodiment, cell based assays can be exploited to analyze a variegated kinase library. For example, a library of expression vectors can be transfected into a cell line which ordinarily synthesizes and secretes kinase. The transfected cells are then
15 cultured such that kinase and a particular mutant are secreted and the effect of expression of the mutant on activity in cell supernatants can be detected, e.g., by any of a number of enzymatic assays. Plasmid DNA can then be recovered from the cells which score for inhibition, or alternatively, potentiation of kinase activity, and the individual clones further characterized.

20 An isolated 14790 protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind kinase using standard techniques for polyclonal and monoclonal antibody preparation. A full-length 14790 protein can be used or, alternatively, the invention provides antigenic peptide fragments of kinase for use as immunogens. The antigenic peptide of kinase comprises at least 8 amino acid residues of
25 the amino acid sequence shown in SEQ ID NO:2 and encompasses an epitope of kinase such that an antibody raised against the peptide forms a specific immune complex with kinase. Preferably, the antigenic peptide comprises at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

30 Preferred epitopes encompassed by the antigenic peptide are regions of kinase that are located on the surface of the protein, e.g., hydrophilic regions.

A kinase immunogen typically is used to prepare antibodies by immunizing a suitable subject, (e.g., rabbit, goat, mouse or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for example, recombinantly expressed
35 14790 protein or a chemically synthesized 14790 polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent. Immunization of a suitable subject with an immunogenic kinase preparation induces a polyclonal anti-kinase antibody response.

Accordingly, another aspect of the invention pertains to anti-kinase antibodies. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen binding site which specifically binds (immunoreacts with) an antigen, such as kinase. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and F(ab')₂ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies that bind kinase. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of kinase. A monoclonal antibody composition thus typically displays a single binding affinity for a particular 14790 protein with which it immunoreacts.

Polyclonal anti-kinase antibodies can be prepared as described above by immunizing a suitable subject with a kinase immunogen. The anti-kinase antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized kinase. If desired, the antibody molecules directed against kinase can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, e.g., when the anti-kinase antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497) (see also, Brown et al. (1981) *J. Immunol.* 127:539-46; Brown et al. (1980) *J. Biol. Chem.* 255:4980-83; Yeh et al. (1976) *Proc. Natl. Acad. Sci. USA* 76:2927-31; and Yeh et al. (1982) *Int. J. Cancer* 29:269-75), the more recent human B cell hybridoma technique (Kozbor et al. (1983) *Immunol Today* 4:72), the EBV-hybridoma technique (Cole et al. (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing monoclonal antibody hybridomas is well known (see generally R. H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); E. A. Lerner (1981) *Yale J. Biol. Med.*, 54:387-402; M. L. Gefter et al. (1977) *Somatic Cell Genet.* 3:231-36). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with a kinase immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds kinase.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating an anti-kinase

monoclonal antibody (see, e.g., G. Galfre et al. (1977) *Nature* 266:55052; Gefter et al. *Somatic Cell Genet.*, cited *supra*; Lerner, *Yale J. Biol. Med.*, cited *supra*; Kenneth, *Monoclonal Antibodies*, cited *supra*). Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods which also would be useful.

5 Typically, the immortal cell line (e.g., a myeloma cell line) is derived from the same mammalian species as the lymphocytes. For example, murine hybridomas can be made by fusing lymphocytes from a mouse immunized with an immunogenic preparation of the present invention with an immortalized mouse cell line. Preferred immortal cell lines are mouse myeloma cell lines that are sensitive to culture medium containing hypoxanthine,
10 aminopterin and thymidine ("HAT medium"). Any of a number of myeloma cell lines can be used as a fusion partner according to standard techniques, e.g., the P3-NS1/1-Ag4-1, P3-x63-Ag8.653 or Sp2/O-Ag14 myeloma lines. These myeloma lines are available from ATCC. Typically, HAT-sensitive mouse myeloma cells are fused to mouse splenocytes
15 selected using HAT medium, which kills unfused and unproductively fused myeloma cells (unfused splenocytes die after several days because they are not transformed). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind kinase, e.g., using a standard ELISA assay.

20 Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal anti-kinase antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (e.g., an antibody phage display library) with kinase to thereby isolate immunoglobulin library members that bind kinase. Kits for generating and screening phage display libraries are commercially available (e.g., the Pharmacia
25 *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *SurfZAP™ Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, Ladner et al. U.S. Patent No. 5,223,409; Kang et al. PCT International Publication No. WO 92/18619; Dower et al. PCT International
30 Publication No. WO 91/17271; Winter et al. PCT International Publication WO 92/20791; Markland et al. PCT International Publication No. WO 92/15679; Breitling et al. PCT International Publication WO 93/01288; McCafferty et al. PCT International Publication No. WO 92/01047; Garrard et al. PCT International Publication No. WO 92/09690; Ladner et al. PCT International Publication No. WO 90/02809; Fuchs et al. (1991) *Bio/Technology*
35 9:1370-1372; Hay et al. (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse et al. (1989) *Science* 246:1275-1281; Griffiths et al. (1993) *EMBO J* 12:725-734; Hawkins et al. (1992) *J. Mol. Biol.* 226:889-896; Clarkson et al. (1991) *Nature* 352:624-628; Gram et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:3576-3580; Garrard et al. (1991) *Bio/Technology* 9:1373-

1377; Hoogenboom et al. (1991) *Nuc. Acid Res.* 19:4133-4137; Barbas et al. (1991) *Proc. Natl. Acad. Sci. USA* 88:7978-7982; and McCafferty et al. *Nature* (1990) 348:552-554.

Additionally, recombinant anti-kinase antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson et al. International Application No. PCT/US86/02269; Akira, et al. European Patent Application 184,187; Taniguchi, M., European Patent Application 171,496; Morrison et al. European Patent Application 173,494; Neuberger et al. PCT International Publication No. WO 86/01533; Cabilly et al. U.S. Patent No. 4,816,567; Cabilly et al. European Patent Application 125,023; Better et al. (1988) *Science* 240:1041-1043; Liu et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu et al. (1987) *J. Immunol.* 139:3521-3526; Sun et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura et al. (1987) *Canc. Res.* 47:999-1005; Wood et al. (1985) *Nature* 314:446-449; and Shaw et al. (1988) *J. Natl. Cancer Inst.* 80:1553-1559; Morrison, S. L. (1985) *Science* 229:1202-1207; Oi et al. (1986) *BioTechniques* 4:214; Winter U.S. Patent 5,225,539; Jones et al. (1986) *Nature* 321:552-525; Verhoeyan et al. (1988) *Science* 239:1534; and Beidler et al. (1988) *J. Immunol.* 141:4053-4060.

An anti-kinase antibody (e.g., monoclonal antibody) can be used to isolate kinase by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-kinase antibody can facilitate the purification of natural kinase from cells and of recombinantly produced kinase expressed in host cells. Moreover, an anti-kinase antibody can be used to detect 14790 protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the 14790 protein. Anti-kinase antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

III. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a 14790 protein (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like. The expression vectors of the invention can be introduced

into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., 14790 proteins, mutant forms of 14790 proteins, fusion proteins, and the like).

The recombinant expression vectors of the invention can be designed for
 5 expression of 14790 proteins in prokaryotic or eukaryotic cells. For example, 14790 proteins can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be
 10 transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein
 15 encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion
 20 moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith, D.B. and Johnson, K.S. (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and
 25 pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Purified fusion proteins can be utilized in kinase activity assays, (e.g., direct assays or competitive assays described in detail below), or to generate antibodies specific for 14790 proteins, for example. In a preferred embodiment, a kinase fusion protein expressed
 30 in a retroviral expression vector of the present invention can be utilized to infect bone marrow cells which are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (e.g., six (6) weeks).

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc
 35 (Amann et al., (1988) *Gene* 69:301-315) and pET 11d (Studier et al., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET

11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

5 One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an
10 expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada et al., (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the kinase expression vector is a yeast expression vector.
15 Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari, et al., (1987) *Embo J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz et al., (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and picZ (Invitrogen Corp, San Diego, CA).

Alternatively, 14790 proteins can be expressed in insect cells using baculovirus
20 expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al. (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian
25 expression vectors include pCDM8 (Seed, B. (1987) *Nature* 329:840) and pMT2PC (Kaufman et al. (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both
30 prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g.,
35 tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al. (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.*

43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji et al. (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund et al. (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α -fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to kinase mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub, H. et al., Antisense RNA as a molecular tool for genetic analysis, *Reviews - Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, a 14790 protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms

"transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (*Molecular Cloning: A Laboratory Manual*, 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding a 14790 protein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (i.e., express) a 14790 protein. Accordingly, the invention further provides methods for producing a 14790 protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding a 14790 protein has been introduced) in a suitable medium such that a 14790 protein is produced. In another embodiment, the method further comprises isolating a 14790 protein from the medium or the host cell.

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which kinase-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous kinase sequences have been introduced into their genome or homologous recombinant animals in which endogenous kinase sequences have been altered. Such animals are useful for studying the function and/or activity of a kinase and for identifying and/or evaluating modulators of kinase activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, and the like. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the

genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing a kinase-encoding nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. The kinase cDNA sequence corresponding to SEQ ID NO:1 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a nonhuman homologue of a human kinase gene, such as a mouse or rat kinase gene, can be used as a transgene. Alternatively, a kinase gene homologue, such as another kinase family member, can be isolated based on hybridization to the kinase cDNA sequences of SEQ ID NO:1, or the coding region thereof, (described further in subsection I above) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to a kinase transgene to direct expression of a 14790 protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder et al., U.S. Patent No. 4,873,191 by Wagner et al. and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of a kinase transgene in its genome and/or expression of kinase mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding a 14790 protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a kinase gene into which a deletion, addition or substitution has been introduced to thereby alter, e.g., functionally disrupt, the kinase gene. The kinase gene can be a human gene (e.g., the SEQ ID NO:1), but more preferably, is a non-human homologue of a human kinase gene (e.g., a cDNA isolated by stringent hybridization with the nucleotide sequence of SEQ ID NO:1). For example, a mouse kinase gene can be used to construct a homologous recombination vector suitable for altering an endogenous kinase gene in the mouse genome. In a preferred embodiment, the vector is designed such that,

upon homologous recombination, the endogenous kinase gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous kinase gene is mutated or otherwise altered but still encodes a functional

5 protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous 14790 protein). In the homologous recombination vector, the altered portion of the kinase gene is flanked at its 5' and 3' ends by additional nucleic acid sequence of the gene to allow for homologous recombination to occur between the exogenous kinase gene carried by the vector and an endogenous kinase gene in an

10 embryonic stem cell. The additional flanking kinase nucleic acid sequence is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see e.g., Thomas, K.R. and Capecchi, M. R. (1987) *Cell* 51:503 for a description of homologous recombination vectors). The vector is introduced into an embryonic stem cell

15 line (e.g., by electroporation) and cells in which the introduced kinase gene has homologously recombined with the endogenous kinase gene are selected (see, e.g., Li, E. et al. (1992) *Cell* 69:915). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (see e.g., Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E.J. Robertson, ed.

20 (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous

25 recombination vectors and homologous recombinant animals are described further in Bradley, A. (1991) *Current Opinion in Biotechnology* 2:823-829 and in PCT International Publication Nos.: WO 90/11354 by Le Mouellec et al.; WO 91/01140 by Smithies et al.; WO 92/0968 by Zijlstra et al.; and WO 93/04169 by Berns et al.

In another embodiment, transgenic non-humans animals can be produced which

30 contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman et al. (1991) *Science*

35 251:1351-1355. If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a

transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. et al. (1997) *Nature* 385:810-813 and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G₀ phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

IV. Pharmaceutical Compositions

The kinase nucleic acid molecules, 14790 proteins, and anti-kinase antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium

hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a 14790 protein or anti-kinase antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or

compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD50 (the dose lethal to 50% of the population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD50/ED50. Compounds which exhibit large therapeutic indices are preferred. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC50 (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U.S. Patent 5,328,470) or by stereotactic injection (see *e.g.*, Chen et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

V. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) predictive medicine (*e.g.*, diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics); and c) methods of treatment (*e.g.*, therapeutic and prophylactic).

006220: 9089T60

The isolated nucleic acid molecules of the invention can be used, for example, to express 14790 protein (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect kinase mRNA (e.g., in a biological sample) or a genetic alteration in a kinase gene, and to modulate kinase activity, as described further below. The 14790 proteins can be used to treat disorders characterized by insufficient or excessive production of a kinase substrate or production of kinase inhibitors. In addition, the 14790 proteins can be used to screen for naturally occurring kinase substrates, to screen for drugs or compounds which modulate kinase activity, as well as to treat disorders characterized by insufficient or excessive production of 14790 protein or production of 14790 protein forms which have decreased or aberrant activity compared to kinase wild type protein. Moreover, the anti-kinase antibodies of the invention can be used to detect and isolate 14790 proteins, regulate the bioavailability of 14790 proteins, and modulate kinase activity.

15 A. Screening Assays:

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, i.e., candidate or test compounds or agents (e.g., peptides, peptidomimetics, small molecules or other drugs) which bind to 14790 proteins, have a stimulatory or inhibitory effect on, for example, kinase expression or kinase activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of a kinase substrate.

In one embodiment, the invention provides assays for screening candidate or test compounds which are substrates of a 14790 protein or polypeptide or biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of a 14790 protein or polypeptide or biologically active portion thereof, e.g., modulate the ability of kinase to interact with its cognate ligand. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) *Anticancer Drug Des.* 12:145).

35 Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann *et al.* (1994). *J. Med. Chem.*

37:2678; Cho et al. (1993) *Science* 261:1303; Carrell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carrell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop et al. (1994) *J. Med. Chem.* 37:1233.

Libraries of compounds may be presented in solution (e.g., Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner USP 5,223,409), spores (Ladner USP '409), plasmids (Cull et al. (1992) *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390); (Devlin (1990) *Science* 249:404-406); (Cwirla et al. (1990) *Proc. Natl. Acad. Sci.* 87:6378-6382); (Felici (1991) *J. Mol. Biol.* 222:301-310); (Ladner *supra.*).

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a kinase target molecule (e.g., a kinase phosphorylation substrate) with a test compound and determining the ability of the test compound to modulate (e.g. stimulate or inhibit) the activity of the kinase target molecule. Determining the ability of the test compound to modulate the activity of a kinase target molecule can be accomplished, for example, by determining the ability of the 14790 protein to bind to or interact with the target molecule, or by determining the ability of the protein to phosphorylate the kinase target molecule.

The ability of the protein to phosphorylate a kinase target molecule can be determined by, for example, an *in vitro* kinase assay. Briefly, a kinase target molecule, e.g., an immunoprecipitated kinase target molecule from a cell line expressing such a molecule, can be incubated with the 14790 protein and radioactive ATP, e.g., [γ - 32 P] ATP, in a buffer containing MgCl₂ and MnCl₂, e.g., 10 mM MgCl₂ and 5 mM MnCl₂. Following the incubation, the immunoprecipitated kinase target molecule can be separated by SDS-polyacrylamide gel electrophoresis under reducing conditions, transferred to a membrane, e.g., a PVDF membrane, and autoradiographed. The appearance of detectable bands on the autoradiograph indicates that the kinase substrate has been phosphorylated. Phosphoaminoacid analysis of the phosphorylated substrate can also be performed in order to determine which residues on the substrate are phosphorylated. Briefly, the radiophosphorylated protein band can be excised from the SDS gel and subjected to partial acid hydrolysis. The products can then be separated by one-dimensional electrophoresis and analyzed on, for example, a phosphoimager and compared to ninhydrin-stained phosphoaminoacid standards.

Determining the ability of the 14790 protein to bind to or interact with a kinase target molecule can be accomplished by determining direct binding. Determining the ability of the 14790 protein to bind to or interact with a target molecule can be accomplished, for example, by coupling the 14790 protein with a radioisotope or enzymatic label such that binding of the protein to a kinase target molecule can be

determined by detecting the labeled protein in a complex. For example, molecules, e.g., proteins, can be labeled with ^{125}I , ^{35}S , ^{14}C , or ^3H , either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, molecules can be enzymatically labeled with, for example, horseradish
5 peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

It is also within the scope of this invention to determine the ability of a compound to modulate the interaction between kinase and its target molecule, without the labeling of any of the interactants. For example, a microphysiometer can be used to detect the
10 interaction of with its target molecule without the labeling of either kinase or the target molecule. McConnell, H. M. et al. (1992) *Science* 257:1906-1912. As used herein, a "microphysiometer" (e.g., Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction
15 between compound and receptor.

In a preferred embodiment, determining the ability of the 14790 protein to bind to or interact with a kinase target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (e.g., intracellular Ca^{2+} ,
20 diacylglycerol, IP_3 , etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising a target-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, e.g., chloramphenicol acetyl transferase), or detecting a target-regulated cellular response.

In yet another embodiment, an assay of the present invention is a cell-free assay in
25 which a protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the 14790 protein or biologically active portion thereof is determined. Binding of the test compound to the 14790 protein can be determined either directly or indirectly as described above. In a preferred embodiment, the assay includes contacting the 14790 protein or biologically active portion thereof with a
30 known compound which binds kinase to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a 14790 protein, wherein determining the ability of the test compound to interact with a 14790 protein comprises determining the ability of the test compound to preferentially bind to kinase or biologically active portion thereof as compared to the known compound.

35 In another embodiment, the assay is a cell-free assay in which a 14790 protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the 14790 protein or biologically active portion thereof is determined. Determining the ability of the test

compound to modulate the activity of a 14790 protein can be accomplished, for example, by determining the ability of the 14790 protein to bind to a kinase target molecule by one of the methods described above for determining direct binding. Determining the ability of the 14790 protein to bind to a kinase target molecule can also be accomplished using a technology such as real-time Biomolecular Interaction Analysis (BIA). Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem.* 63:2338-2345 and Szabo et al. (1995) *Curr. Opin. Struct. Biol.* 5:699-705. As used herein, "BIA" is a technology for studying biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore). Changes in the optical phenomenon of surface plasmon resonance (SPR) can be used as an indication of real-time reactions between biological molecules.

In an alternative embodiment, determining the ability of the test compound to modulate the activity of a 14790 protein can be accomplished by determining the ability of the 14790 protein to further modulate the activity of a kinase target molecule (e.g., a kinase mediated signal transduction pathway component). For example, the activity of the effector molecule on an appropriate target can be determined, or the binding of the effector to an appropriate target can be determined as previously described.

In yet another embodiment, the cell-free assay involves contacting a 14790 protein or biologically active portion thereof with a known compound which binds the 14790 protein to form an assay mixture, contacting the assay mixture with a test compound, and
20 determining the ability of the test compound to interact with the 14790 protein, wherein determining the ability of the test compound to interact with the 14790 protein comprises determining the ability of the 14790 protein to preferentially bind to or modulate the activity of a kinase target molecule.

The cell-free assays of the present invention are amenable to use of both soluble and/or membrane-bound forms of proteins (e.g., 14790 proteins or biologically active portions thereof, or receptors to which kinase binds). In the case of cell-free assays in which a membrane-bound form a 14790 protein is used (e.g., a cell surface kinase receptor) it may be desirable to utilize a solubilizing agent such that the membrane-bound form of the 14790 protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)_n, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl=N,N-dimethyl-3-ammonio-1-propane sulfonate.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either kinase or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as

to accommodate automation of the assay. Binding of kinase a test compound to a protein, or interaction of a 14790 protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates, test tubes, and micro-
5 centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/kinase fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the test
10 compound or the test compound and either the non-adsorbed target protein or 14790 protein, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtitre plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described
15 above. Alternatively, the complexes can be dissociated from the matrix, and the level of kinase binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either a 14790 protein or a kinase target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated
20 14790 protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with 14790 protein or target molecules but which do not interfere with binding of the 14790 protein to its target
25 molecule can be derivatized to the wells of the plate, and unbound target or 14790 protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodection of complexes using antibodies reactive with the 14790 protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity
30 associated with the 14790 protein or target molecule.

In another embodiment, modulators of kinase expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of kinase mRNA or protein in the cell is determined. The level of expression of kinase mRNA or protein in the presence of the candidate compound is compared to the level of expression of kinase
35 mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of kinase expression based on this comparison. For example, when expression of kinase mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate

compound is identified as a stimulator of kinase mRNA or protein expression. Alternatively, when expression of kinase mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of kinase mRNA or protein expression.

- 5 The level of kinase mRNA or protein expression in the cells can be determined by methods described herein for detecting kinase mRNA or protein.

In yet another aspect of the invention, the 14790 proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos et al. (1993) *Cell* 72:223-232; Madura et al. (1993) *J. Biol. Chem.* 268:12046-12054; Bartel et al. (1993) *Biotechniques* 14:920-924; Iwabuchi et al. (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with kinase ("kinase-binding proteins" or "kinase-bp") and are involved in kinase activity. Such kinase-binding proteins are also likely to be involved in the propagation of signals by the 14790 proteins or kinase targets as, for example, downstream elements of a kinase-mediated signaling pathway. Alternatively, such kinase-binding proteins are likely to be kinase inhibitors.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a 14790 protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a kinase dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the 14790 protein which interacts with the protein.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (e.g., a kinase modulating agent, an antisense kinase nucleic acid molecule, a kinase-specific antibody, or a kinase-binding partner) can be used in an animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal model to determine the mechanism of action of such an agent. Furthermore, this invention

pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

B. Detection Assays

5 Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic
10 identification of a biological sample. These applications are described in the subsections below.

1. Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this
15 sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the kinase nucleotide sequences, described herein, can be used to map the location of the kinase genes on a chromosome. The mapping of the kinase sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

20 Briefly, kinase genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the kinase nucleotide sequences. Computer analysis of the kinase sequences can be used to predict primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human
25 chromosomes. Only those hybrids containing the human gene corresponding to the kinase sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (e.g., human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse
30 chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but human cells can, the one human chromosome that contains the gene encoding the needed enzyme, will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse
35 chromosomes, allowing easy mapping of individual genes to specific human chromosomes. (D'Eustachio P. et al. (1983) *Science* 220:919-924). Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the kinase nucleotide sequences to design oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map a 90,
1p, or 1v kinase sequence to its chromosome include *in situ* hybridization (described in Fan, Y. et al. (1990) *Proc. Natl. Acad. Sci. USA*, 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical such as colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma et al., Human Chromosomes: A Manual of Basic Techniques (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland, J. et al. (1987) *Nature*, 325:783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison

of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

2. Tissue Typing

The kinase sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique which determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the kinase nucleotide sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The kinase nucleotide sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:1 can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as nucleotides 1-4950 in SEQ ID NO:1 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from nucleotide sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

3. Use of Partial Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology.

Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NO:1 are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the kinase nucleotide sequences or portions thereof, e.g., fragments derived from the noncoding regions of SEQ ID NO:1 having a length of at least 20 bases, preferably at least 30 bases.

The kinase nucleotide sequences described herein can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, e.g., brain tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such kinase probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these reagents, e.g., kinase primers or probes can be used to screen tissue culture for contamination (i.e. screen for the presence of a mixture of different types of cells in a culture).

C. Predictive Medicine:

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining 14790 protein and/or nucleic acid expression as well as kinase activity, in the context of a biological sample (e.g., blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant kinase expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with 14790 protein, nucleic acid expression or activity. For example, mutations in a kinase gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with 14790 protein, nucleic acid expression or activity.

Another aspect of the invention pertains to monitoring the influence of agents (e.g., drugs, compounds) on the expression or activity of kinase in clinical trials.

These and other agents are described in further detail in the following sections.

1. Diagnostic Assays

An exemplary method for detecting the presence or absence of 14790 protein or nucleic acid in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting 14790 protein or nucleic acid (e.g., mRNA, genomic DNA) that encodes 14790 protein such that the presence of 14790 protein or nucleic acid is detected in the biological sample. A preferred agent for detecting mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to mRNA or genomic DNA. The nucleic acid probe can be, for example, a human nucleic acid, such as the nucleic acid of SEQ ID NO:1 or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to kinase mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

A preferred agent for detecting protein is an antibody capable of binding to protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or F(ab')₂) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or

antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include
5 tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect kinase mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of kinase mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of 14790
10 protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of kinase genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of 14790 protein include introducing into a subject a labeled anti-kinase antibody. For example, the antibody can be labeled with a radioactive marker whose
15 presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains 14790 protein molecules from the test subject. Alternatively, the biological sample can contain kinase mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a serum sample isolated by conventional means from a subject.

20 In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting 14790 protein, mRNA, or genomic DNA, such that the presence of 14790 protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of 14790 protein, mRNA or genomic DNA in the control sample
25 with the presence of 14790 protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of kinase in a biological sample. For example, the kit can comprise a labeled compound or agent capable of detecting 14790 protein or mRNA in a biological sample; means for determining the amount of kinase in the sample; and means for comparing the amount of kinase in the
30 sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect 14790 protein or nucleic acid.

2. Prognostic Assays

35 The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant kinase expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject

having or at risk of developing a disorder associated with 14790 protein, nucleic acid expression or activity. Thus, the present invention provides a method for identifying a disease or disorder associated with aberrant kinase expression or activity in which a test sample is obtained from a subject and 14790 protein or nucleic acid (e.g., mRNA, genomic DNA) is detected, wherein the presence of 14790 protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant kinase expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (e.g., serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant kinase expression or activity. Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant kinase expression or activity in which a test sample is obtained and 14790 protein or nucleic acid expression or activity is detected (e.g., wherein the abundance of 14790 protein or nucleic acid expression or activity is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant kinase expression or activity).

The methods of the invention can also be used to detect genetic alterations in a kinase gene, thereby determining if a subject with the altered gene is at risk for a disorder associated with the kinase gene. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic alteration characterized by at least one of an alteration affecting the integrity of a gene encoding a kinase-protein, or the mis-expression of the kinase gene. For example, such genetic alterations can be detected by ascertaining the existence of at least one of 1) a deletion of one or more nucleotides from a kinase gene; 2) an addition of one or more nucleotides to a kinase gene; 3) a substitution of one or more nucleotides of a kinase gene, 4) a chromosomal rearrangement of a kinase gene; 5) an alteration in the level of a messenger RNA transcript of a kinase gene, 6) aberrant modification of a kinase gene, such as of the methylation pattern of the genomic DNA, 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a kinase gene, 8) a non-wild type level of a kinase-protein, 9) allelic loss of a gene, and 10) inappropriate post-translational modification of a kinase-protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting alterations in a kinase gene. A preferred biological sample is a tissue or serum sample isolated by conventional means from a subject, e.g., a hepatic tissue sample.

In certain embodiments, detection of the alteration involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g., U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran et al. (1988) *Science* 241:1077-1080; and Nakazawa et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:360-364), the latter of which can be particularly useful for detecting point mutations in the kinase-gene (see Abravaya et al. (1995) *Nucleic Acids Res.* 23:675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a kinase gene under conditions such that hybridization and amplification of the kinase-gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli, J.C. et al., (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh, D.Y. et al., (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi, P.M. et al. (1988) *Bio-Technology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a kinase gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in kinase can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin, M.T. et al. (1996) *Human Mutation* 7: 244-255; Kozal, M.J. et al. (1996) *Nature Medicine* 2: 753-759). For example, genetic mutations in kinase can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, M.T. *et al. supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a

sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the kinase gene and detect mutations by comparing the sequence of the kinase sample with the corresponding wild-type (control) sequence.

Examples of sequencing reactions include those based on techniques developed by Maxam and Gilbert ((1977) *Proc. Natl. Acad. Sci. USA* 74:560) or Sanger ((1977) *Proc. Natl. Acad. Sci. USA* 74:5463). It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen et al. (1996) *Adv. Chromatogr.* 36:127-162; and Griffin et al. (1993) *Appl. Biochem. Biotechnol.* 38:147-159).

Other methods for detecting mutations in the kinase gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers et al. (1985) *Science* 230:1242). In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes formed by hybridizing (labeled) RNA or DNA containing the wild-type kinase sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, for example, Cotton *et al.* (1988) *Proc. Natl Acad Sci USA* 85:4397; Saleeba et al. (1992) *Methods Enzymol.* 217:286-295. In a preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in kinase cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells

cleaves T at G/T mismatches (Hsu *et al.* (1994) *Carcinogenesis* 15:1657-1662).

According to an exemplary embodiment, a probe based on kinase a sequence, e.g., a wild-type kinase sequence, is hybridized to a cDNA or other DNA product from a test cell(s).

The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if

5 any, can be detected from electrophoresis protocols or the like. See, for example, U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in kinase genes. For example, single strand conformation

polymorphism (SSCP) may be used to detect differences in electrophoretic mobility

10 between mutant and wild type nucleic acids (Orita *et al.* (1989) *Proc Natl. Acad. Sci USA*: 86:2766, see also Cotton (1993) *Mutat Res* 285:125-144; and Hayashi (1992) *Genet Anal*

Tech Appl 9:73-79). Single-stranded DNA fragments of sample and control kinase nucleic acids will be denatured and allowed to renature. The secondary structure of single-

stranded nucleic acids varies according to sequence, the resulting alteration in

15 electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the

20 basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet* 7:5).

In yet another embodiment the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.* (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not

25 completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys Chem* 265:12753).

Examples of other techniques for detecting point mutations include, but are not

30 limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki *et al.* (1986) *Nature* 324:163); Saiki *et al.* (1989) *Proc. Natl Acad. Sci USA* 86:6230). Such allele specific

35 oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al.* (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner *et al.* (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini *et al.* (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a kinase gene.

Furthermore, any cell type or tissue in which kinase is expressed may be utilized in the prognostic assays described herein.

3. Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs or compounds) on the expression or activity of a 14790 protein can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase kinase gene expression, protein levels, or upregulate kinase activity, can be monitored in clinical trials of subjects exhibiting decreased kinase gene expression, protein levels, or downregulated kinase activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease kinase gene expression, protein levels, or downregulate kinase activity, can be monitored in clinical trials of subjects exhibiting increased kinase gene expression, protein levels, or upregulated activity. In such clinical trials, the expression or activity of a kinase gene, and preferably, other genes that have been implicated in a disorder can be used as a "read out" or markers of the phenotype of a particular cell.

For example, and not by way of limitation, genes, including kinase, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) which modulates kinase activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on a kinase associated disorder, for

example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of kinase and other genes implicated in the kinase associated disorder, respectively. The levels of gene expression (i.e., a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of kinase or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during treatment of the individual with the agent.

In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a 14790 protein, mRNA, or genomic DNA in the pre-administration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the 14790 protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the 14790 protein, mRNA, or genomic DNA in the pre-administration sample with the 14790 protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of kinase to higher levels than detected, i.e., to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of kinase to lower levels than detected, i.e. to decrease the effectiveness of the agent. According to such an embodiment, kinase expression or activity may be used as an indicator of the effectiveness of an agent, even in the absence of an observable phenotypic response.

C. Methods of Treatment:

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant kinase expression or activity. With regards to both prophylactic and therapeutic methods of treatment, such treatments may be specifically tailored or modified, based on knowledge obtained from the field of pharmacogenomics. "Pharmacogenomics", as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. More specifically, the term refers the study of how a patient's genes determine

his or her response to a drug (e.g., a patient's "drug response phenotype", or "drug response genotype".) Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with either the 14790 molecules of the present invention or kinase modulators according to that individual's drug response

5 genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the treatment and to avoid treatment of patients who will experience toxic drug-related side effects.

1. Prophylactic Methods

10 In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant kinase expression or activity, by administering to the subject a or an agent which modulates kinase expression or at least one kinase activity. Subjects at risk for a disease which is caused or contributed to by aberrant kinase expression or activity can be identified by, for example, any or a combination of diagnostic

15 or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of kinase aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of kinase aberrancy, for example, a kinase, kinase agonist or kinase antagonist agent can be used for treating the subject. The appropriate agent can be determined based

20 on screening assays described herein.

2. Therapeutic Methods

Another aspect of the invention pertains to methods of modulating kinase expression or activity for therapeutic purposes. Accordingly, in an exemplary

25 embodiment, the modulatory method of the invention involves contacting a cell with a kinase or agent that modulates one or more of the activities of 14790 protein activity associated with the cell. An agent that modulates 14790 protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring target molecule of a 14790 protein (e.g., a kinase phosphorylation substrate), a kinase antibody, a kinase

30 agonist or antagonist, a peptidomimetic of a kinase agonist or antagonist, or other small molecule. In one embodiment, the agent stimulates one or more kinase activities. Examples of such stimulatory agents include active 14790 protein and a nucleic acid molecule encoding kinase that has been introduced into the cell. In another embodiment, the agent inhibits one or more kinase activities. Examples of such inhibitory agents include

35 antisense kinase nucleic acid molecules, anti-kinase antibodies, and kinase inhibitors. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or

disorder characterized by aberrant expression or activity of a 14790 protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or downregulates) kinase expression or activity. In another embodiment, the method involves administering a 14790 protein or nucleic acid molecule as therapy to compensate for reduced or aberrant kinase expression or activity.

Stimulation of kinase activity is desirable in situations in which kinase is abnormally downregulated and/or in which increased kinase activity is likely to have a beneficial effect. For example, stimulation of kinase activity is desirable in situations in which a kinase is downregulated and/or in which increased kinase activity is likely to have a beneficial effect. Likewise, inhibition of kinase activity is desirable in situations in which kinase is abnormally upregulated and/or in which decreased kinase activity is likely to have a beneficial effect.

3. Pharmacogenomics

The 14790 molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on kinase activity (e.g., kinase gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (e.g, infection with HBV) associated with aberrant kinase activity. In conjunction with such treatment, pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, a physician or clinician may consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer a 14790 molecule or kinase modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with a 14790 molecule or kinase modulator.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, for example, Eichelbaum, M. et al. (1996) *Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 and Linder, M.W. et al. (1997) *Clin. Chem.* 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main

clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

One pharmacogenomics approach to identifying genes that predict drug response, known as "a genome-wide association", relies primarily on a high-resolution map of the human genome consisting of already known gene-related markers (e.g., a "bi-allelic" gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants.) Such a high-resolution genetic map can be compared to a map of the genome of each of a statistically significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. Alternatively, such a high resolution map can be generated from a combination of some ten-million known single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a "SNP" is a common alteration that occurs in a single nucleotide base in a stretch of DNA. For example, a SNP may occur once per every 1000 bases of DNA. A SNP may be involved in a disease process, however, the vast majority may not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular pattern of SNPs in their individual genome. In such a manner, treatment regimens can be tailored to groups of genetically similar individuals, taking into account traits that may be common among such genetically similar individuals.

Alternatively, a method termed the "candidate gene approach", can be utilized to identify genes that predict a drug response. According to this method, if a gene that encodes a drug target is known (e.g., a 14790 protein or kinase receptor of the present invention), all common variants of that gene can be fairly easily identified in the population and it can be determined if having one version of the gene versus another is associated with a particular drug response.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (e.g., N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as

demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Alternatively, a method termed the "gene expression profiling", can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (e.g., a 14790 molecule or kinase modulator of the present invention) can give an indication whether gene pathways related to toxicity have been turned on.

Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment an individual. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a 14790 molecule or kinase modulator, such as a modulator identified by one of the exemplary screening assays described herein.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application are incorporated herein by reference.

EXAMPLES

EXAMPLE 1: IDENTIFICATION AND CHARACTERIZATION OF HUMAN cDNA

Isolation of the human cDNA

The invention is based, at least in part, on the discovery of a human gene encoding a novel member of the kinase family. The human kinase family members were isolated from cDNA. A cDNA library was prepared therefrom using art known methods (described in, for example, *Molecular Cloning A Laboratory Manual*, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989). Positive clones were isolated following comparison to homologs in public protein databases, including a comparison with known kinases and/or examination of the sequence for protein motifs of kinases.

The sequences of the positive clones were determined and found to contain open reading frames. The nucleotide sequence encoding the human 14790 protein is shown in Figure 1 and is set forth as SEQ ID NO:1. The protein encoded by this nucleic acid comprises about 1650 amino acids and has the amino acid sequence shown in Figure 1 and

set forth as SEQ ID NO:2. The coding region (open reading frame) of SEQ ID NO:1 is shown in Figure 1 as the portion of the nucleotide sequence corresponding to the amino acid sequence of SEQ ID NO:2. The clone comprising the entire coding region of human kinase was deposited with the American Type Culture Collection (ATCC®), 10801 University Boulevard, Manassas, VA 20110-2209, on _____, 1998, and assigned Accession No. _____.

Analysis of human 14790 molecule

A BLASTN 1.4.9 search, using a score of 100 and a word length of 12 (Altschul et al. (1990) *J. Mol. Biol.* 215:403) of the nucleotide sequence of human kinase revealed that kinase is similar to the human protein kinase HPK-1 coding sequence (Accession No. V23831). This nucleic acid molecule is approximately 70% identical to kinase, over nucleotides 388-1214.

Tissue Distribution of Kinase mRNA

This Example describes the tissue distribution of kinase mRNA, as determined by TaqMan RT-PCR and *in situ* hybridization.

14790 mRNA was found to be expressed in human skeletal muscle, brain and liver. TaqMan RT-PCR analysis revealed that 14790 mRNA was found to be upregulated in liver cells which were infected with HBV. Moreover, mRNA expression of 14790 was found to be restricted to hepatocytes of HBV infected livers as seen by *in situ* hybridization. 14790 mRNA was also found to be upregulated in HepG2.2.15 cells (HBV positive) compared to HepG2 parent cells (HBV negative). When HepG2.2.15 cells were treated with anti-HBV drug treatment, the upregulation of 14790 mRNA was eliminated. Thus indicating that a modulator of 14790 activity or mRNA may be used to treat infection by HBV.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

What is claimed:

1. An isolated nucleic acid molecule selected from the group consisting of:
 - 5 (a) a nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:1; and
 - (b) a nucleic acid molecule comprising the coding region of the nucleotide sequence set forth in SEQ ID NO:1.
- 10 2. An isolated nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:2.
3. An isolated nucleic acid molecule selected from the group consisting of:
 - 15 (a) a nucleic acid molecule comprising the nucleotide sequence contained in the plasmid deposited with ATCC® as Accession Number _____.
4. An isolated nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule comprising SEQ
 - 20 ID NO:1 or the coding region of SEQ ID NO:1 under stringent conditions.
5. An isolated nucleic acid molecule selected from the group consisting of:
 - a) a nucleic acid molecule comprising a nucleotide sequence which is
 - 25 at least 60% homologous to the nucleotide sequence of SEQ ID NO:1 or the coding region of SEQ ID NO:1;
 - b) a nucleic acid molecule comprising a fragment of at least 200 nucleotides of a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1 or the coding region of SEQ ID NO:1, or a complement thereof;
 - 30 c) a nucleic acid molecule which encodes a polypeptide comprising an amino acid sequence at least about 60% homologous to the amino acid sequence of SEQ ID NO:2; and
 - d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the fragment comprises at
 - 35 least 15 contiguous amino acid residues of the amino acid sequence of SEQ ID NO:2.
6. An isolated nucleic acid molecule which hybridizes to the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 under stringent conditions.

7. An isolated nucleic acid molecule comprising a nucleotide sequence which is complementary to the nucleotide sequence of the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5.

8. An isolated nucleic acid molecule comprising the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5, and a nucleotide sequence encoding a heterologous polypeptide.

9. A vector comprising the nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5.

10. The vector of claim 9, which is an expression vector.

11. A host cell transfected with the vector of claim 9.

12. A method of producing a polypeptide comprising culturing a host cell transfected with the vector of claim 9 in an appropriate culture medium to, thereby, produce the polypeptide.

13. An isolated polypeptide selected from the group consisting of:

a) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2;

b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, or the coding region of SEQ ID NO:1 under stringent conditions;

c) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 60 % homologous to a nucleic acid comprising the nucleotide sequence of SEQ ID NO:1, or the coding region of SEQ ID NO:1; and

d) a polypeptide comprising an amino acid sequence which is at least 60% homologous to the amino acid sequence of SEQ ID NO:2.

14. The isolated polypeptide of claim 13 comprising the amino acid sequence of SEQ ID NO:2.

15. The polypeptide of claim 13, further comprising heterologous amino acid sequences.

16. An antibody which selectively binds to a polypeptide of claim 13.

17. A method for detecting the presence of a polypeptide of claim 13 in a sample comprising:

a) contacting the sample with a compound which selectively binds to the polypeptide; and

b) determining whether the compound binds to the polypeptide in the sample to thereby detect the presence of a polypeptide of claim 13 in the sample.

18. The method of claim 17, wherein the compound which binds to the polypeptide is an antibody.

19. A kit comprising a compound which selectively binds to a polypeptide of claim 13 and instructions for use.

20. A method for detecting the presence of a nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 in a sample comprising:

a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes to the nucleic acid molecule; and

b) determining whether the nucleic acid probe or primer binds to a nucleic acid molecule in the sample to thereby detect the presence of a nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 in the sample.

21. The method of claim 20, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

22. A kit comprising a compound which selectively hybridizes to a nucleic acid molecule of any one of claims 1, 2, 3, 4, or 5 and instructions for use.

23. A method for identifying a compound which binds to a polypeptide of claim 13 comprising:

a) contacting the polypeptide, or a cell expressing the polypeptide with a test compound; and

b) determining whether the polypeptide binds to the test compound.

24. The method of claim 23, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

- a) detection of binding by direct detection of test compound/polypeptide binding;
- 5 b) detection of binding using a competition binding assay; and
- c) detection of binding using an assay for activity.

25. A method for modulating the activity of a polypeptide of claim 13 comprising contacting the polypeptide or a cell expressing the polypeptide with a
10 compound which binds to the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

26. A method for identifying a compound which modulates the activity of a polypeptide of claim 13 comprising:

- 15 a) contacting a polypeptide of claim 13 with a test compound; and
- b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.

20 27. A method for treating a subject having a hepatic disorder characterized by aberrant 14790 kinase polypeptide activity or aberrant 14790 nucleic acid expression comprising administering to the subject a 14790 kinase modulator, thereby treating said subject having a hepatic associated disorder.

25 28. The method of claim 27 wherein the 14790 kinase modulator is a small molecule.

29. The method of claim 27, wherein the disorder is liver fibrosis.

30. The method of claim 27, wherein the disorder is infection with HBV.

Abstract of the Disclosure

1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 2560 2561 2562 2563 2564 2565 2566 2567 2568 2569 2570 2571 2572 2573 2574 2575 2576 2577 2578 2579 2580 2581 2582 2583 2584 2585 2586 2587 2588 2589 2590 2591 2592 2593 2594 2595 2596 2597 2598 2599 2600 2601 2602 2603 2604 2605 2606 2607 2608 2609 2610 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620 2621 2622 2623 2624 2625 2626 2627 2628 2629 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640 2641 2642 2643 2644 2645 2646 2647 2648 2649 2650 2651 2652 2653 2654 2655 2656 2657 2658 2659 2660 2661 2662 2663 2664 2665 2666 2667 2668 2669 2670 2671 2672 2673 2674 2675 2676 2677 2678 2679 2680 2681 2682 2683 2684 2685 2686 2687 2688 2689 2690 2691 2692 2693 2694 2695 2696 2697 2698 2699 2700 2701 2702 2703 2704 2705 2706 2707 2708 2709 2710 2711 2712 2713 2714 2715 2716 2717 2718 2719 2720 2721 2722 2723 2724 2725 2726 2727 2728 2729 2730 2731 2732 2733 2734 2735 2736 2737 2738 2739 2740 2741 2742 2743 2744 2745 2746 2747 2748 2749 2750 2751 2752 2753 2754 2755 2756 2757 2758 2759 2760 2761 2762 2763 2764 2765 2766 2767 2768 2769 2770 2771 2772 2773 2774 2775 2776 2777 2778 2779 2780 2781 2782 2783 2784 2785 2786 2787 2788 2789 2790 2791 2792 2793 2794 2795 2796 2797 2798 2799 2800 2801 2802 2803 2804 2805 2806 2807 2

SEQUENCE ID NO:1

FIG. 1a

AAT CTC AAA GAG CAA GAC GAC TCC ATC GTG GTG GAC ATT TTA GTG GAG CAC ATT AGT GGG	1152
V S L A A H L S H S G P I P V H Q L R R	404
GTC TCT CTT GCT GCA CAC CTG AGC CAC TCA GGC CCC ATC CCT GTG CAT CAG CTT CGC AGG	1212
Y T A Q L L S G L D Y L H S N S V V H K	424
TAC ACA GCT CAG CTC CTG TCA GGC CTT GAT TAT CTG CAC AGC AAT TCT GTG GTG CAT AAG	1272
V L S A S N V L V D A E G T V K I T D Y	444
GTC CTG AGT GCA TCT AAT GTC TTG GTG GAT GCA GAA GGC ACC GTC AAG ATT ACG GAC TAT	1332
S I S K R L A D I C K E D V F E Q T R V	464
AGC ATT TCT AAG CGC CTC GCA GAC ATT TGC AAG GAG GAT GTG TTT GAG CAA ACC CGA GTT	1392
R F S D N A L P Y K T G K K G D V W R L	484
CGT TTT AGT GAC AAT GCT CTG CCT TAT AAA ACG GGG AAG AAA GGA GAT GTT TGG CGT CTT	1452
G L L L L S L S Q G Q E C G E Y P V T I	504
GGC CTT CTG CTG CTG TCC CTC AGC CAA GGA CAG GAA TGT GGA GAG TAC CCT GTG ACC ATC	1512
P S D L P A D F Q D F L K K C V C L D D	524
CCT AGT GAC TTA CCA GCT GAC TTT CAA GAT TTT CTA AAG AAA TGT GTG TGC TTG GAT GAC	1572
K E R W S P Q Q L L K H S F I N P Q P K	544
AAG GAA AGA TGG AGT CCC CAG CAG TTG TTG AAA CAC AGC TTT ATA AAT CCC CAG CCA AAA	1632
M P L V E Q S P E D S G G Q D Y V E T V	564
ATG CCT CTA GTG GAA CAA AGT CCT GAA GAT TCT GGA GGA CAA GAT TAT GTT GAG ACT GTT	1692
I P S N R L P S A A F F S E T Q R Q F S	584
ATT CCT AGC AAC CGG CTA CCC AGT GCT GCC TTC TTT AGT GAG ACA CAG AGA CAG TTT TCC	1752
R Y F I E F E E L Q L L G K G A F G A V	604
CGA TAC TTC ATT GAG TTT GAA GAA TTA CAA CTT CTT GGT AAA GGA GCT TTT GGA GCT GTC	1812
I K V Q N K L D G C C Y A V K R I P I N	624
ATC AAG GTG CAG AAC AAG TTG GAC GGC TGC TGC TAC GCA GTG AAG CGC ATC CCC ATC AAC	1872
P A S R Q F R R I K G E V T L L S R L H	644
CCG GCC AGC CGG CAG TTC CGC AGG ATC AAG GGC GAA GTG ACA CTG CTG TCA CGG CTG CAC	1932
H E N I V R Y Y N A W I E R H E R P A G	664
CAT GAG AAC ATT GTG CGC TAC TAC AAC GCC TGG ATC GAG CGG CAC GAG CGG CCG GCG GGA	1992
P G T P P P D S G P L A K D D R A A R G	684
CCG GGG ACG CCG CCC CCG GAC TCC GGG CCC CTG GCC AAG GAT GAC CGA GCT GCA CGC GGG	2052
Q P A S D T D G L D S V E A A A P P P I	704
CAG CCG GCG AGC GAC ACA GAC GGC CTG GAC AGC GTA GAG GCC GCC GCG CCG CCA CCC ATC	2112
L S S S V E W S T S G E R S A S A R F P	724
CTC AGC AGC TCG GTG GAG TGG AGC ACT TCG GGC GAG CGC TCG GCC AGT GCC CGT TTC CCC	2172
A T G P G S S D D E D D D E D E H G G V	744
GCC ACC GGC CCG GGC TCC AGC GAT GAC GAG GAC GAC GAC GAG GAC GAG CAC GGT GGC GTC	2232
F S Q S F L P A S D S E S D I I F D N E	764
TTC TCC CAG TCC TTC CTG CCT GCT TCA GAT TCT GAA AGT GAT ATT ATC TTT GAC AAT GAA	2292
D E N S K S Q N Q D E D C N E K N G C H	784
GAT GAG AAC AGT AAA AGT CAG AAT CAG GAT GAA GAT TGC AAT GAA AAG AAT GGC TGC CAT	2352

FIG. 1b

E S E P S V T T E A V H Y L Y I Q M E Y	804
GAA AGT GAG CCA TCA GTG ACG ACT GAG GCT GTG CAC TAC CTA TAC ATC CAG ATG GAG TAC	2412
C E K S T L R D T I D Q G L Y R D T V R	824
TGT GAG AAG AGC ACT TTA CGA GAC ACC ATT GAC CAG GGA CTG TAT CGA GAC ACC GTC AGA	2472
L W R L F R E I L D G L A Y I H E K G M	844
CTC TGG AGG CTT TTT CGA GAG ATT CTG GAT GGA TTA GCT TAT ATC CAT GAG AAA GGA ATG	2532
I H R D L K P V N I F L D S D D H V K I	864
ATT CAC CGG GAT TTG AAG CCT GTC AAC ATT TTT TTG GAT TCT GAT GAC CAT GTG AAA ATA	2592
G D F G L A T D H L A F S A D S K Q D D	884
GGT GAT TTT GGT TTG GCG ACA GAC CAT CTA GCC TTT TCT GCT GAC AGC AAA CAA GAC GAT	2652
Q T G D L I K S D P S G H L T G M V G T	904
CAG ACA GGA GAC TTG ATT AAG TCA GAC CCT TCA GGT CAC TTA ACT GGG ATG GTT GGC ACT	2712
A L Y V S P E V Q G S T K S A Y N Q K V	924
GCT CTC TAT GTA AGC CCA GAG GTC CAA GGA AGC ACC AAA TCT GCA TAC AAC CAG AAA GTG	2772
D L F S L G I I F F E M S Y H P M V T A	944
GAT CTC TTC AGC CTG GGA ATT ATC TTC TTT GAG ATG TCC TAT CAC CCC ATG GTC ACG GCT	2832
S E R I F V L N Q L R D P T S P K F P E	964
TCA GAA AGG ATC TTT GTT CTC AAC CAA CTC AGA GAT CCC ACT TCG CCT AAG TTT CCA GAA	2892
D F D D G E H A K Q K S V I S W L L N H	984
GAC TTT GAC GAT GGA GAG CAT GCA AAG CAG AAA TCA GTC ATC TCC TGG CTG TTG AAC CAC	2952
D P A K R P T A T E L L K S E L L P P P	1004
GAT CCA GCA AAA CGG CCC ACA GCC ACA GAA CTG CTC AAG AGT GAG CTG CTG CCC CCA CCC	3012
Q M E E S E L H E V L H H T L T N V D G	1024
CAG ATG GAG GAG TCA GAG CTG CAT GAA GTG CTG CAC CAC ACG CTG ACC AAC GTG GAT GGG	3072
K A Y R T M M A Q I F S Q R I S P A I D	1044
AAG GCC TAC CGC ACC ATG ATG GCC CAG ATC TTC TCG CAG CGC ATC TCC CCT GCC ATC GAT	3132
Y T Y D S D I L K G N F S I R T A K M Q	1064
TAC ACC TAT GAC AGC GAC ATA CTG AAG GGC AAC TTC TCA ATC CGT ACA GCC AAG ATG CAG	3192
Q H V C E T I I R I F K R H G A V Q L C	1084
CAG CAT GTG TGT GAA ACC ATC ATC CGC ATC TTT AAA AGA CAT GGA GCT GTT CAG TTG TGT	3252
T P L L L P R N R Q I Y E H N E A A L F	1104
ACT CCA CTA CTG CTT CCC CGA AAC AGA CAA ATA TAT GAG CAC AAC GAA GCT GCC CTA TTC	3312
M D H S G M L V M L P F D L R I P F A R	1124
ATG GAC CAC AGC GGG ATG CTG GTG ATG CTT CCT TTT GAC CTG CGG ATC CCT TTT GCA AGA	3372
Y V A R N N I L N L K R Y C I E R V F R	1144
TAT GTG GCA AGA AAT AAT ATA TTG AAT TTA AAA CGA TAC TGC ATA GAA CGT GTG TTC AGG	3432
P R K L D R F H P K E L L E C A F D I V	1164
CCG CGC AAG TTA GAT CGA TTT CAT CCC AAA GAA CTT CTG GAG TGT GCA TTT GAT ATT GTC	3492
T S T T N S F L P T A E I I Y T I Y E I	1184
ACT TCT ACC ACC AAC AGC TTT CTG CCC ACT GCT GAA ATT ATC TAC ACT ATC TAT GAA ATC	3552

FIG. 1c

I	Q	E	F	P	A	L	Q	E	R	N	Y	S	I	Y	L	N	H	T	M	1204
ATC	CAA	GAG	TTT	CCA	GCA	CTT	CAG	GAA	AGA	AAT	TAC	AGT	ATT	TAT	TTG	AAC	CAT	ACC	ATG	3612
L	L	K	A	I	L	L	H	C	G	I	P	E	D	K	L	S	Q	V	Y	1224
TTA	TTG	AAA	GCA	ATA	CTC	TTA	CAC	TGT	GGG	ATC	CCA	GAA	GAT	AAA	CTC	AGT	CAA	GTC	TAC	3672
I	I	L	Y	D	A	V	T	E	K	L	T	R	R	E	V	E	A	K	F	1244
ATT	ATT	CTG	TAT	GAT	GCT	GTG	ACA	GAG	AAG	CTG	ACG	AGG	AGA	GAA	GTG	GAA	GCT	AAA	TTT	3732
C	N	L	S	L	S	S	N	S	L	C	R	L	Y	K	F	I	E	Q	K	1264
TGT	AAT	CTG	TCT	TTG	TCT	TCT	AAT	AGT	CTG	TGT	CGA	CTC	TAC	AAG	TTT	ATT	GAA	CAG	AAG	3792
G	D	L	Q	D	L	M	P	T	I	N	S	L	I	K	Q	K	T	G	I	1284
GGA	GAT	TTG	CAA	GAT	CTT	ATG	CCA	ACA	ATA	AAT	TCA	TTA	ATA	AAA	CAG	AAA	ACA	GGT	ATT	3852
A	Q	L	V	K	Y	G	L	K	D	L	E	E	V	V	G	L	L	K	K	1304
GCA	CAG	TTG	GTG	AAG	TAT	GGC	TTA	AAA	GAC	CTA	GAG	GAG	GTT	GTT	GGA	CTG	TTG	AAG	AAA	3912
L	G	I	K	L	Q	V	L	I	N	L	G	L	V	Y	K	V	Q	Q	H	1324
CTC	GGC	ATC	AAG	TTA	CAG	GTC	TTG	ATC	AAT	TTG	GGC	TTG	GTT	TAC	AAG	GTG	CAG	CAG	CAC	3972
N	G	I	I	F	Q	F	V	A	F	I	K	R	R	Q	R	A	V	P	E	1344
AAT	GGA	ATC	ATC	TTC	CAG	TTT	GTG	GCT	TTC	ATC	AAA	CGA	AGG	CAA	AGG	GCT	GTA	CCT	GAA	4032
I	L	A	A	G	G	R	Y	D	L	L	I	P	Q	F	R	G	P	Q	A	1364
ATC	CTC	GCA	GCT	GGA	GGC	AGA	TAT	GAC	CTG	CTG	ATT	CCC	CAG	TTT	AGA	GGG	CCA	CAA	GCT	4092
L	G	P	V	P	T	A	I	G	V	S	I	A	I	D	K	I	S	A	A	1384
CTG	GGG	CCA	GTT	CCC	ACT	GCC	ATT	GGG	GTC	AGC	ATA	GCT	ATA	GAC	AAG	ATA	TCT	GCT	GCT	4152
V	L	N	M	E	E	S	V	T	I	S	S	C	D	L	L	V	V	S	V	1404
GTC	CTC	AAC	ATG	GAG	GAA	TCT	GTT	ACA	ATA	AGC	TCT	TGT	GAC	CTC	CTG	GTT	GTA	AGT	GTT	4212
G	Q	M	S	M	S	R	A	I	N	L	T	Q	K	L	W	T	A	G	I	1424
GGT	CAG	ATG	TCT	ATG	TCC	AGG	GCC	ATC	AAC	CTA	ACC	CAG	AAA	CTC	TGG	ACA	GCA	GGC	ATC	4272
T	A	E	I	M	Y	D	W	S	Q	S	Q	E	E	L	Q	E	Y	C	R	1444
ACA	GCA	GAA	ATC	ATG	TAC	GAC	TGG	TCA	CAG	TCC	CAA	GAG	GAA	TTA	CAA	GAG	TAC	TGC	AGA	4332
H	H	E	I	T	Y	V	A	L	V	S	D	K	E	G	S	H	V	K	V	1464
GAT	CAT	GAA	ATC	ACC	TAT	GTG	GCC	CTT	GTC	TCG	GAT	AAA	GAA	GGA	AGC	CAT	GTC	AAG	GTT	4392
K	S	F	E	K	E	R	Q	T	E	K	R	V	L	E	T	E	L	V	D	1484
AAG	TCT	TTC	GAG	AAG	GAA	AGG	CAG	ACA	GAG	AAG	CGT	GTG	CTG	GAG	ACT	GAA	CTT	GTG	GAC	4452
H	V	L	Q	K	L	R	T	K	V	T	D	E	R	N	G	R	E	A	S	1504
CAT	GTA	CTG	CAG	AAA	CTG	AGG	ACT	AAA	GTC	ACT	GAT	GAA	AGG	AAT	GGC	AGA	GAA	GCT	TCC	4512
D	N	L	A	V	Q	N	L	K	G	S	F	S	N	A	S	G	L	F	E	1524
GAT	AAT	CTT	GCA	GTG	CAA	AAT	CTG	AAG	GGG	TCA	TTT	TCT	AAT	GCT	TCA	GGT	TTG	TTT	GAA	4572
I	H	G	A	T	V	V	P	I	V	S	V	L	A	P	E	K	L	S	A	1544
ATC	CAT	GGA	GCA	ACA	GTG	GTT	CCC	ATT	GTG	AGT	GTG	CTA	GCC	CCG	GAG	AAG	CTG	TCA	GCC	4632
S	T	R	R	R	Y	E	T	Q	V	Q	T	R	L	Q	T	S	L	A	N	1564
AGC	ACT	AGG	AGG	CGC	TAT	GAA	ACT	CAG	GTA	CAA	ACT	CGA	CTT	CAG	ACC	TCC	CTT	GCC	AAC	4692
L	H	Q	K	S	S	E	I	E	I	L	A	V	D	L	P	K	E	T	I	1584
TTA	CAT	CAG	AAA	AGC	AGT	GAA	ATT	GAA	ATT	CTG	GCT	GTG	GAT	CTA	CCC	AAA	GAA	ACA	ATA	4752
L	Q	F	L	S	L	E	W	D	A	D	E	Q	A	F	N	T	T	V	K	1604

FIG. 1d

TTA CAG TTT TTA TCA TTA GAG TGG GAT GCT GAT GAA CAG GCA TTT AAC ACA ACT GTG AAG	4812
Q L L S R L P K Q R Y L K L V C D E I Y	1624
CAG CTG CTG TCA CGC CTG CCA AAG CAA AGA TAC CTC AAA TTA GTC TGT GAT GAA ATT TAT	4872
N I K V E K K V S V L F L Y S Y R D D Y	1644
AAC ATC AAA GTA GAA AAA AAG GTG TCT GTG CTA TTT CTG TAC AGC TAT AGA GAT GAC TAC	4932
Y R I L F *	1650
TAC AGA ATC TTA TTT TAA	4950

CCCTAAAGAAGCTGTCGTTAACCTCATTTCAAACAGACAGAGGCTTTATACTGGAATAATGGAATGTTGTACATTCATCATA
ATTTTAAAATTAAATTCTAAGAAGAGGCTGGGTGCAGTGGCTCACACCTTTAATCCCAGCACTTTGGGAAGCCAAGGCAG
GAAGACTGCTTTGAAACCAGGAGTTTGAGACCAGCCTGAGCAACAAAGCAAGACCCCATCTCTATAAAAACTAAAAAAT
TAGTTGGGCATGGTGGCACATGCCTGTAGTCCCAGCTACTCCAGAGGCTGAGATGGATCATCTGAGCCTCAGGAGGTTG
AGGCTGCAGTGAGCTGTGACTGCGCCACTGCACTCCAGTCTGGGACAACAGAGCAAGACCCCTGTCTTAAAAAAAAAAG
AAAAAAAAAATTTTTTTCTAAGAAGCTGTCTACAAAGTTGAGCTTTGTTAGTTTTCATGTGTAATATATTATAAATT
TATCTTTTGGGATATAATAAATGCTTTCATATACCTGCA

FIG. 1e

006220 90851560

CLUSTAL W (1.74) multiple sequence alignment

```

6065914_eIF2kinase_man  -----
14790                   MAGGRGAPGRGRDEPPESYPQRQDHELQALEAIYGADFQDLRPDACPVK
6066585_eIF2kinase_mouse MAGGRGASGRGRAEPQESYSQRQDHELQALEAIYGSDFQDLRPDARGVR

6065914_eIF2kinase_man  -----
14790                   EPPEINLVLYPQGLTGEEVYVKVDLRVKCPPTYPDVVPEIELKNAKGLSN
6066585_eIF2kinase_mouse EPPEINLVLYPQGLAGEEVYVQVELQVKCPPTYPDVVPEIELKNAKGLSN

6065914_eIF2kinase_man  -----
14790                   ESVNLLKSRLEELAKKHCGEVMIFELAYHVQSFLSEHNKPPPKSFHEEML
6066585_eIF2kinase_mouse ESVNLLKSHLEELAKKQCGEVMIFELAHVQSFLSEHNKPPPKSFHEEML

6065914_eIF2kinase_man  -----
14790                   ERRAQEEQQRLLEAKRKEEQEQREILHEIQRRKEEIKEEKKRKEMAKQER
6066585_eIF2kinase_mouse ERQAQEKQQRLLLEARRKEEQEQREILHEIQRRKEEIKEEKKRKEMAKQER

6065914_eIF2kinase_man  -----
14790                   LEIASLSNQDHTSKKDPGGHRTAAIILHGGSPDFVGNGKHRANSSGRSRRE
6066585_eIF2kinase_mouse LEITSLTNQDYASKRDPAGHRAAAIILHGGSPDFVGNGKARTYSSGRSRRE

6065914_eIF2kinase_man  -----
14790                   RQYSVCNSEDSPGSCCEILYFNMGSPDQLMVHKGKCTIGSDEQLGKLVYNAL
6066585_eIF2kinase_mouse RQYSVCSGEPSPGSCDILHFSVGSPDQLMVHKGRCVGSDEQLGKVYNAL

6065914_eIF2kinase_man  -----
14790                   ETATGGFVLLYEWVLQWQKKMGPFLLTSQEKEKIDKCKKQIQGTETEFNSL
6066585_eIF2kinase_mouse ETATGSFVLLHEWVLQWQK-MGPCLTSQEKEKIDKCKRQIQGAETEFSSL

6065914_eIF2kinase_man  -----
14790                   VKLSHPNVVRYLAMNLKEQDDSI VVDILVEHISGVSLAAHLSHSGPIPVH
6066585_eIF2kinase_mouse VKLSHPNIVRYFAMNSREEEDSIVIDILAEHVSGISLATHLSHSGPVPAPH

6065914_eIF2kinase_man  -----
14790                   QLRRYTAQLLSGLDYLSNSVVKVLSASNVLVDAEGTVKITDYSISKRL
6066585_eIF2kinase_mouse QLRKYTAQLLAGLDYLSNSVVKVLSASSVLVDAEGTVKITDYSISKRL

6065914_eIF2kinase_man  -----
14790                   ADICKEDVFEQTRVRFSDNALPYKTGKKGDVWRLGLLLLSLSQGQECGEY
6066585_eIF2kinase_mouse ADICKEDVFEQARVRFSDSALPYKTGKKGDVWRLGLLLLSLSQGQECGEY

6065914_eIF2kinase_man  -----
14790                   PVTIPSDLPADFDQFLKKCVCLDDKERWSPQQLLKHSFINPQPKMPLVEQ
6066585_eIF2kinase_mouse PVTIPSDLPADFDQFLKKCVCLDDKERWSPQQLLKHSFINPQPKLPLVEQ

6065914_eIF2kinase_man  -----
14790                   SPEDSGGQDYVETVIPS NRLPSA AFFSETQRQFSRYFIEFEELQLLGKGA
6066585_eIF2kinase_mouse SPEDSGGQDYIETVIPS NQLPSA AFFSETQKQFSRYFIEFEELQLLGKGA

6065914_eIF2kinase_man  -----
14790                   FGAVIKVQNKL DGCCYAVKRIPINPASRQFRRIKGEVTLLSRLHHENIVR
6066585_eIF2kinase_mouse FGAVIKVQNKL DGCCYAVKRIPINPASRHFRRIKGEVTLLSRLHHENIVR

6065914_eIF2kinase_man  -----
14790                   YNAWIERHERPAGPGT PPPDSGPLAKDDRAARGQPASDTDGLDSVEAAA
6066585_eIF2kinase_mouse YNAWIERHERPAVPGT PPPDCTPQAQDSPATCGKTS GDTEELGSVEAAA

```

FIG. 2a

6065914_eIF2kinase_man 14790 6066585_eIF2kinase_mouse	----- PPPISSSSVEWSTSGERSASARFPATGPGSSDDEDDDEHGGVFSQSFL PPPISSSSVEWSTSAERSTSTRFPVTGQDSSSDEED-EDERDGVFSQSFL
6065914_eIF2kinase_man 14790 6066585_eIF2kinase_mouse	----- PASDSSESDIIFDNEDENSKSQNQDEDCNEKNGCHESEPSVTTEAVHYLYI PASDSDSDIIFDNEDENSKSQNQDEDCNQKDGSHIEPSVTAEAVHYLYI
6065914_eIF2kinase_man 14790 6066585_eIF2kinase_mouse	----- QMEYCEKSTLRDTIDQGLYRDTVRLWRLFREILDGLAYIHEKGMHRDLK QMEYCEKSTLRDTIDQGLFRDTSRLWRLFREILDGLAYIHEKGMHRDLK
6065914_eIF2kinase_man 14790 6066585_eIF2kinase_mouse	----- PVNIFLDSDDHVKIGDFGLATDHLAFSADSKQDDQTGD-LIKSDPSGHLT PVNIFLDSDDHVKIGDFGLATDHLAFTAEGKQDDQAGDGVIKSDPSGHLT
6065914_eIF2kinase_man 14790 6066585_eIF2kinase_mouse	----- GMVGTALYVSPEVQGSTKSAYNQKVDLFLSGIIFFEMSYHPMTASERIF GMVGTALYVSPEVQGSTKSAYNQKVDLFLSGIIFFEMSYHPMTASERIF
6065914_eIF2kinase_man 14790 6066585_eIF2kinase_mouse	----- VLNQLRDPSTSPKFPDDFDDGEHAKQKSVISWLLNHDPKRPTATELLKSE VLNQLRDPSTSPKFPDDFDDGEHTKQKSVISWLLNHDPKRPTAMELLKSE
6065914_eIF2kinase_man 14790 6066585_eIF2kinase_mouse	----- LLPPPQMEESSELHEVLHHTLTNVDGKAYRTMMAQIFSQRISPADITYDS LLPPPQMEESSELHEVLHHTLANIDGKAYRTMMSQIFCQHISPADITYDS
6065914_eIF2kinase_man 14790 6066585_eIF2kinase_mouse	----- DILKGNFSIRTAKMQQHVCETIIRIFKRHGAVQLCTPLLLPRNRQIYEHN DILKGNFLIRTAKIQQLVCETIVRVFKRHGAVQLCTPLLLPRNRQIYEHN
6065914_eIF2kinase_man 14790 6066585_eIF2kinase_mouse	----- --ALFMDHSGMLVMLPFDLRIPFARYVARNNILILKRYCIERVFRPRKLD EAALFMDHSGMLVMLPFDLRIPFARYVARNNILNLKRYCIERVFRPRKLD EAALFMDHSGMLVMLPFDLRVPFARYVARNNILNLKRYCIERVFRPRKLD *****:*****:*****:*****
6065914_eIF2kinase_man 14790 6066585_eIF2kinase_mouse	RFHPKELLECAFDIVTSTTNSFLPTAEIIYTIYEIIQEFPALQERNYSIY RFHPKELLECAFDIVTSTTNSFLPTAEIIYTIYEIIQEFPALQERNYSIY RFHPKELLECAFDIVTSTTNSSLPTAETIYTIYEIIQEFPALQERNYSIY *****:*****:*****:*****
6065914_eIF2kinase_man 14790 6066585_eIF2kinase_mouse	LNHTMLLKAILLHCGIPEDKLSQVYIILYDAVTEKLTRREVEAKFCNLSL LNHTMLLKAILLHCGIPEDKLSQVYIILYDAVTEKLTRREVEAKFCNLSL LNHTMLLKAILLHCGIPEDKLSQVYIILYDAVTEKLTRREVEAKFCNLSL *****:*****:*****:*****
6065914_eIF2kinase_man 14790 6066585_eIF2kinase_mouse	SSNSLCRLYKFIEQKGDLDLDMPTINSLIKQKTGIAQLVKYGLKDLEEVV SSNSLCRLYKFIEQKGDLDLDMPTINSLIKQKTGIAQLVKYGLKDLEEVV SSNSLCRLYKFIEQKGDLDLTPTINSLIKQKTGVAQLVKYSLKDLEDVV *****:*****:*****:*****
6065914_eIF2kinase_man 14790 6066585_eIF2kinase_mouse	GLLKKLGIKLQVLINLGLVYKVQQHNGIIFQFVAFIKRRQRAVPEILAAG GLLKKLGIKLQVLINLGLVYKVQQHNGIIFQFVAFIKRRQRAVPEILAAG GLLKKLGKLVQSINLGLVYKVQQHTGIIFQFLAFSKRRQRVVPEILAAG *****:*****:*****:*****
6065914_eIF2kinase_man 14790 6066585_eIF2kinase_mouse	GRYDLLIPQFRGPQALGPVPTAIGVSIADKISAAVLNMEESVTISSCDL GRYDLLIPQFRGPQALGPVPTAIGVSIADKISAAVLNMEESVTISSCDL GRYDLLIPKFRGPQTVGPVPTAVGVSIADKIFAVVLNMEEPVTVSSCDL *****:*****:*****:*****
6065914_eIF2kinase_man	LVVSVGQMSMSRAINLTQKLWTAGITAEIMYDWSQSQEELQEYCRHHEIT

FIG. 2b

14790	LVVSVGQMMSRAINLTQKLWTAGITAEIMYDWSQSQEELQEYCRHHEIT
6066585_eIF2kinase_mouse	LVVSVGQMMSRAINLTQKLWTAGITAEIMYDWSQSQEELQEYCRHHEIT

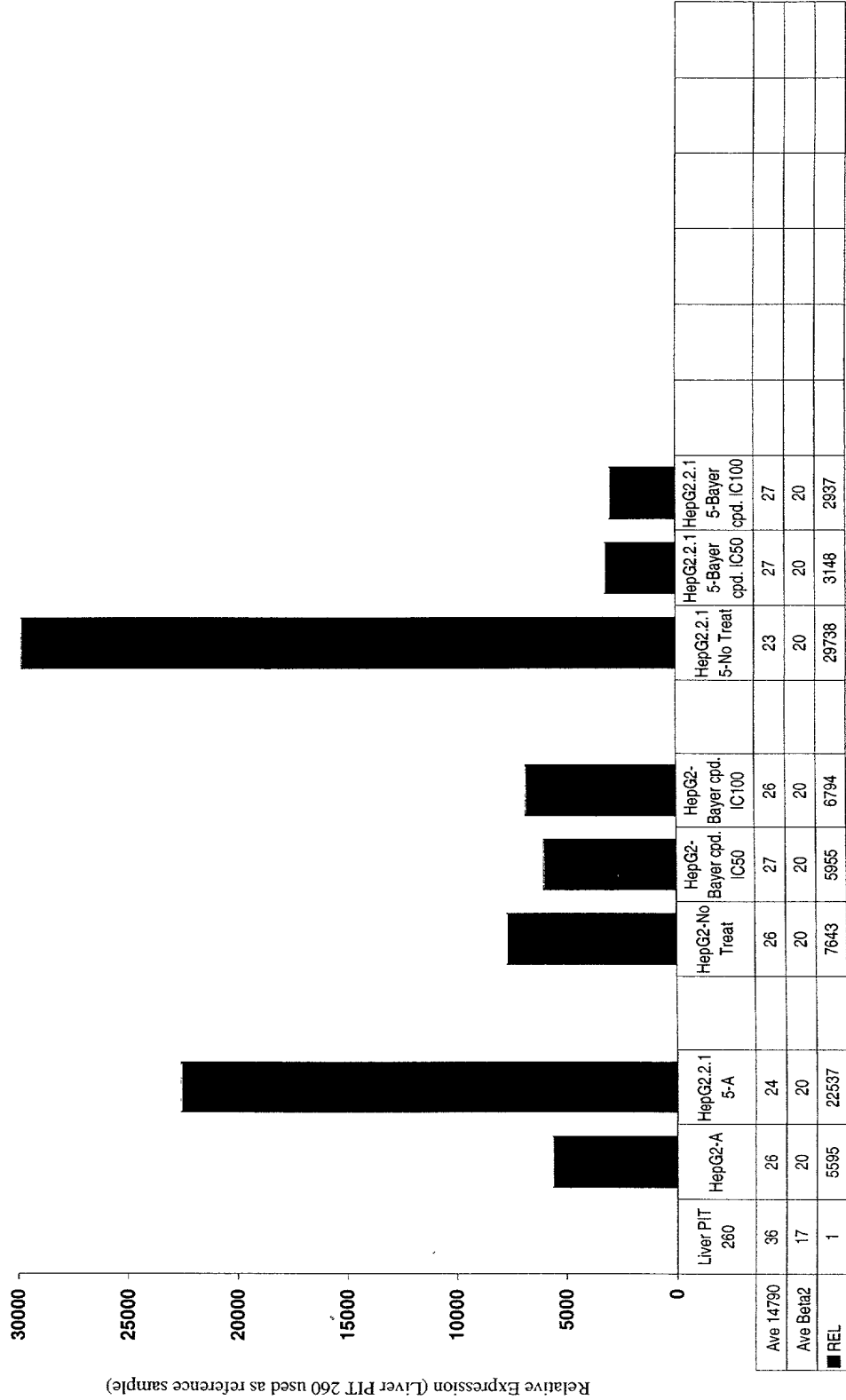
6065914_eIF2kinase_man	YVALVSDKEGSHVKVKSFEKERQTEKRVLETETELVDHVLQKLRTKVTDERN
14790	YVALVSDKEGSHVKVKSFEKERQTEKRVLETETELVDHVLQKLRTKVTDERN
6066585_eIF2kinase_mouse	YVALVSDKEGSHVKVKSFEKERQTEKRVLESDELVDHVMQKLRTKVGDERN
*****:;*****:***** ****	
6065914_eIF2kinase_man	GREASDNLAVQNLKGSFNASGLFEIHGATVVPPIVSVLAPEKLSASTRRR
14790	GREASDNLAVQNLKGSFNASGLFEIHGATVVPPIVSVLAPEKLSASTRRR
6066585_eIF2kinase_mouse	FRDASDNLAVQTLKGSFNASGLFEIHGTTVVPNVIVLAPEKLSASTRRR
*:*****.*****.*****:**** * *****	
6065914_eIF2kinase_man	YETQVQTRLQTSANLHQSSEIEILAVDLPKETILQFLSLEWDAEQAF
14790	YETQVQTRLQTSANLHQSSEIEILAVDLPKETILQFLSLEWDAEQAF
6066585_eIF2kinase_mouse	HEIQVQTRLQTTLANLHQSSEIEILAVDLPKETILQFLSLEWDAEQAF
:* *****:*****	
6065914_eIF2kinase_man	NTTVKQLLSRLPKQRYLKLVCDEIYNIKVEKKVSVFLYSYRDDYYRILF
14790	NTTVKQLLSRLPKQRYLKLVCDEIYNIKVEKKVSVFLYSYRDDYYRILF
6066585_eIF2kinase_mouse	NTTVKQLLSRLPKQRYLKLVCDEIYNIKVEKKVSVFLYSYRDDYYRILF

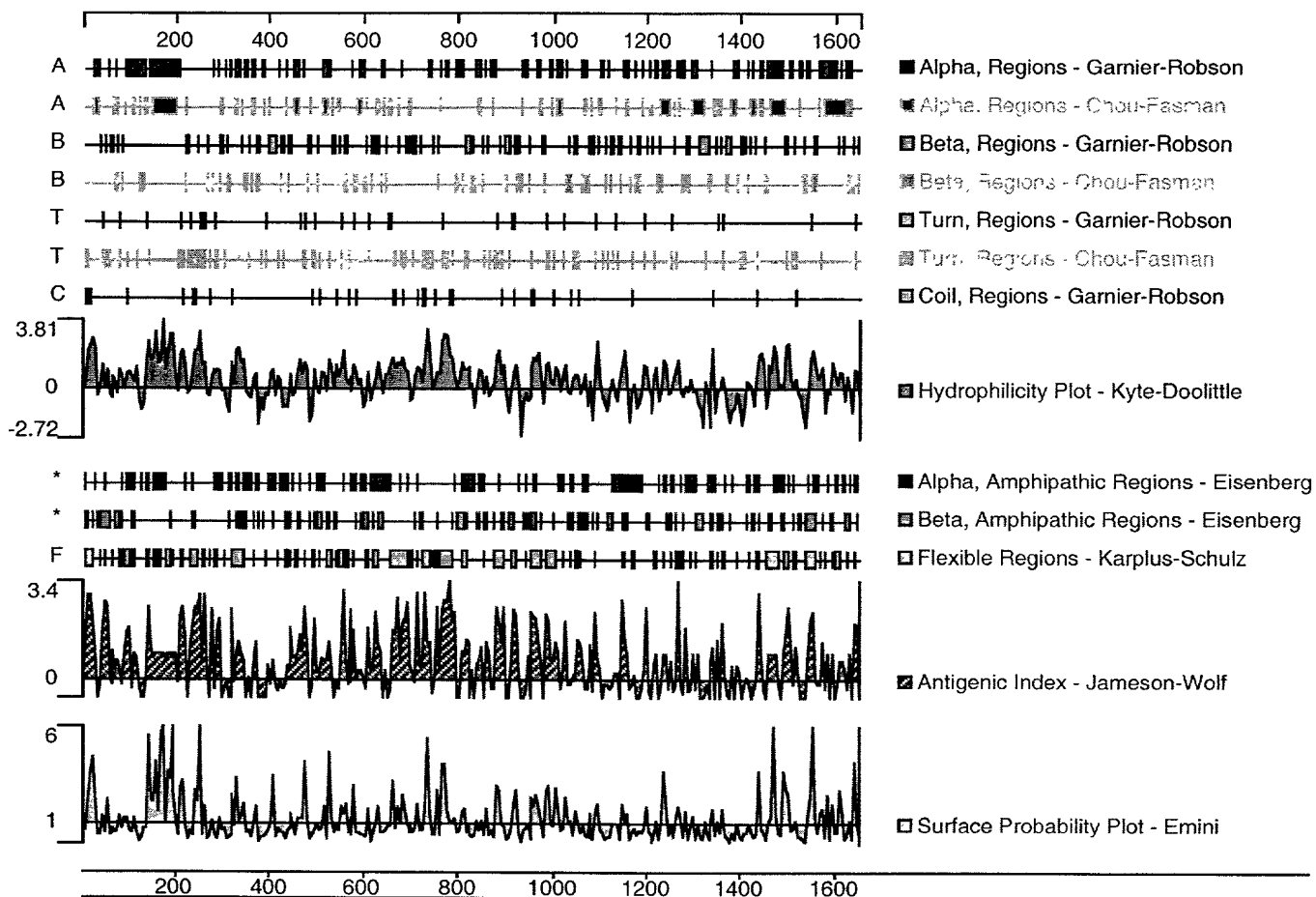
FIG. 2c

Ser/thr Kinase 14790 Expression on HBV+ Liver



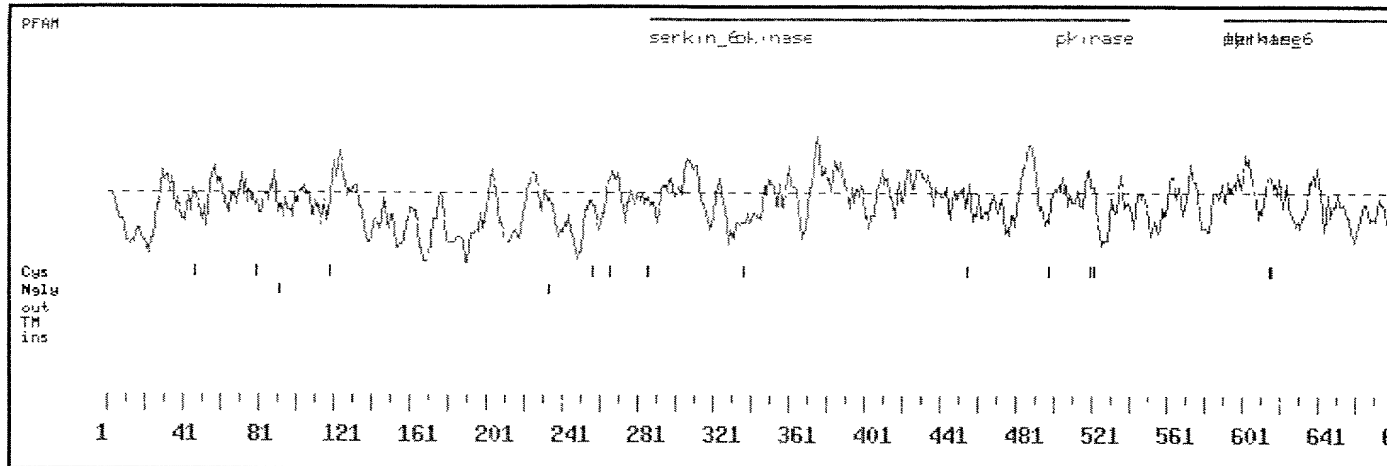
14790 Expression on HepG2.2.15 (HBV-expressing cell line)





[Back to orfanal.cgi](#)

Analysis of 14790 (1649 aa)



```
>14790
MAGGRGAPGRGRDEPPESYPQRQDHELQALEAIYGADFQDLRPDACPVPKEPPEINLVLY
PQGLTGEEVYVVKVDLKVCKPPTYPDVVPEIELKNAKGLSNESVNLLKSRLEELAKKHCGE
VHIFELAYHVQSPFLSEHNKPPPKSFHEEMLERRAQEEQQRLEAKRKEEQREILHEIQ
RKKEEKKEKKRREMAKQERLEIASLSNQDHTSKKDPGGHRTAAILHGGSPDFVGNKGR
ASSGSRRRERQYVSCNSEDSPGSCILYFNMGSPDQLMVHKGKICIGSDEQLGKLVYNAL
EATGGFVLLYEVLQWQKKMGPFLLTSQEKEKIDKCKKQIQGTETEFNSLVKLSHPNVVR
YLAAMLKEQDDSIIVDILVEHISGVSLAHLSHSGPIPVHQLRRYTAQLLSGLDYLSNS
VYHKVLSASNVLVDAEGTVKITDYSISKRLADICKEDVFEQTRVRFSDNALPYKTGKGD
LRLGLLLLLSLSQGQECGEYPTIPSDLPADFQDFLKKCVCLDDKERWSPQQLKHSFIN
PQPKMLPLVQSPEDSGGQDYVETVIPSRLPSAFAFFSETQRFQSRFYIEFEELQLLGKA
FGAVIKVQNKLDGCCYAVKRIPINPASRQFRRIKGEVTLSSLRHENIVRYNNAWIERHE
RAGPGTPPPDSGPLAKDDRAARGQPASDTDGLDSVEAAAPPPILSSSVWSTSGERSAS
RFPATGPGSSDDEDDDEHGGVFSQSFLPASDSSEDIIFDNEDESKSQNQDEDCNEK
GCHSESPSVTTEAVHYLYIQMEYCEKSTLRDTIDQGLYRDTVRLWRLFREILDGLAYIH
FKGMIHRDLKPVNIFLDSDDHVKIGDFGLATDHLAFSADSKQDDQTDGLIKSDPSGHLTG
MYGTALYVSPVEQGSKTSAYNQKVDLFLSLGIIFFEMSYHPMVTASERIFVLNQLRDPSTP
KFPEDFDDGEHAKQKSVISWLLNHDPAKRPTATELLKSELLPPQMESELHEVLHHTLT
VVDGKAYRTMAQIFSQRISPAIDYTYDSILKGNFSIRTAKMQOHVCETIIRIFKRHGA
VOLCTPLLLPRNRQIYEHNEAALFMDHSGMLVMLPFDLRIPFARYVARNNILNLKRYCIE
RYFRPRKLDLRFHPKELLECAFDIVTSTTNSFLPTAEIITYTYEIIQEFPALQERNYSIYL
NMTMLLKAILLHCGIPEDKLSQVYIILYDAVTEKLTREVEAKFCNLSLSSNSLCRLYKF
EQKGLDQLDMPTINSLIKQKTGIAQLVKYGLKDLEEVVGLLKKLGKQLVNLGLVYK
VQOHNGIIFQFVAFIKRRQRAVPEILAAGGRYDLIPQFRGPALGPVPTAIGVSIADK
TAAVNLMEESVTISSCDLLVSVGQMSMSRAINLTQKLWTAGITAEIMYDWSQSQEELQ
EACRHHIITYVALVSDKEGSHVKVKSFEKERQTEKRVLETLDVHVLQKLRKVTDERNG
REASDNLAQNQLKGSFNASGLFEIHGATVVPVSVLAPEKLSASTRRRYETQVQTRLQT
SLANLHKSSEIEILAVDLPKETILQFLSLEWDADEQAFNTTVKQLLSRLPKQRYLKLV
DEIYNIKVEKKVSVFLYSYRDDYYRILF
```

PSORT Prediction of Protein Localization

MITDISC: discrimination of mitochondrial targeting seq

R content:	3	Hyd Moment(75):	7.37
Hyd Moment(95):	6.02	G content:	5
D/E content:	2	S/T content:	0
Score:	-6.53		

Gavel: prediction of cleavage sites for mitochondrial preseq

R-2 motif at 20 GRG|RD

NUCDISC: discrimination of nuclear localization signals

pat4: KRRK (5) at 190
 pat4: RPRK (4) at 1144
 pat7: none
 bipartite: RRAQEEQQRLLLEAKRKE at 152
 bipartite: KRIPINPASRQFRRIKG at 619
 content of basic residues: 12.1%
 NLS Score: 1.08

ER Membrane Retention Signals:

FIG. 7a

http://tango.mpi.com/seqanal/orfanal/14790_16890.html

XXRR-like motif in the N-terminus: AGGR

none

Final Results (k = 9/23):

56.5 %: nuclear
30.4 %: cytoplasmic
4.3 %: vacuolar
4.3 %: mitochondrial
4.3 %: vesicles of secretory system

prediction for 14790 is **nuc** (k=23)

Start	End	Feature	Seq
144	204	coiled coil	FHEEMLERRA...AKQERLEIAS

Signal Peptide Predictions for 14790

Method	Predict	Score	Mat@
SignalP (eukaryote)	NO		

Note: amino-terminal 70aa used for signal peptide prediction

Transmembrane Segments Predicted by MEMSAT

Start	End	Orient	Score
1522	1538	ins-->out	0.8

14790
MGGGRGAPGRGRDEPPESYPQRQDHELQALEAIYGADFQDLRPDACGPVKEPPEINLVLY
PQGLTGEEVYVKVLDLRVKCPPTYPDVVPEIELKNAKGLSNESVNLLKSRLEELAKKHCGE
VPIFELAYHVQSFLSEHNKPPPKSFHEEMLERRAQEEQRLLEAKRKEEQEQRILHEIQ
EKEETKEKKRREMAKQERLEIASLSNQDHTSKKDPGGHRTAAILHGGSPDFVGNKHR
ANSSGRSRRERQYSVCNSEDSPGSCCELLYFNMGSPDQLMVHKGKCIQGSDEQLGLVYNAL
ETATGGFVLLYEWVLQWKKMGPFLLTSQEKEKIDCKKKQIQGTETEFNSLVKLSHPNVVR
YLAANLKEQDDSIIVDILVEHISGVSLAAHLSHSGPIPVHQLRRRYTAQLLSGLDYLSNS
YVHKVLSASNVLVDAEGTVKITDYSISKRLADICKEDVPEQTRVRFSDNALPYKTGKKGD
MRLGLLLLSLQGGQECGEYPVTIPSDLPADFQDFLKKCVCLDDKERWSPQQLKHSFIN
PQPKMPLVEQSPEDSGGQDYVETVIPSRLPSAAFFSETQRFQSRFYIEFEELQLLGKA
FGAVIKVQNKLDGCCYAVKRIPINPASRQFRRIKGEVTLLSRLHENIVRYNAWIERHE
EAGPGTPPPDSGPLAKDDRAARGQPASDTDGLDSVEAAAPPPILSSSVWSTSGERSAS
AFPPATGPGSSDDEDDDEHGGVFSQSFLPASDESIDIIFDNEDENSKSNQDEDCNEK
NGCHESEPSVTTEAVHYLYIQMEYCEKSTLRDTIDQGLYRDTVRLWRLFREILDGLAYIH
EFGMIHRDLKPVNIFLSDDHVKIGDFGLATDHLAFSADSKQDDQTGDLIKSDPSGHLTG
MVGTAlyVSPVQGSTKSAYNQKVDLFSLGIIFFEMSYHPMVTASERIFVLNQLRDPSTP
KPPEDFDGGEHAKQKSVISWLLNHDPAKRPTATELLKSELLPPPMEESELHEVLHHTLT
NVDGKAYRTMMAQIFSQRISPAIDYTYDSILKGNFSIRTAKMQQHVCTIIRIFKRRHGA
VQLCTPLLLPRNRQIYEHNEAALFMDHSGMLVMLPFDLRIFFARYVARNNILNLKRYCIE
RVFRPRKLDRFHPKELLECAFDIVTSTTNSFLPTAEIITYIYIEIIEFFPALQERNYSIYL
NHTMLLKAILLHCGIPEDKLSQVYIILYDAVTEKLTRREVEAKFCNLSSLSSNLCRLYKF
IEQKGDLDQLMPTINSLIKQKTGIAQLVKYGLKLEEVVGLLKKLGIKLQVLINLGLVYK
VQQHNGIIFQFVAFIKRRQRAVPEILAAGGRYDLLIPQFRGPQALGPVPTAIGVSTAIDK
ISAAVLNMEESTVISSCDLLVSVGQMSMSRAINLTQKLWTAGITAEIMYDWSQSQEELQ
EYCRHHEITYVALVSDKEGSHVKVKSFEKERQTEKRVLETETELVDHVLQKLRTKVTDERNG
READNLAVQNLKGSFSNASGLFEIHGATVVPVSVLAPEKLSASTRRRYETQVQTRLQT
SLANLHQKSSEIEILAVDLPKETILQFLSLEWDADEQAFNTTVKQLLSRLPKQRYLKLVC
DEIYNIKVEKKVSVLFLYSYRDDYYRILF

Prosite Pattern Matches for 14790

Prosite version: Release 12.2 of February 1995

>E000001|PDOC00001|ASN_GLYCOSYLATION N-glycosylation site.

Query: 100 NESV 103
Query: 242 NSSG 245

FIG. 7b

Query: 1055 NFSI 1058
 Query: 1195 NYSI 1198
 Query: 1201 NHTM 1204
 Query: 1246 NLSL 1249
 Query: 1414 NLTQ 1417
 Query: 1518 NASG 1521
 Query: 1600 NTTV 1603

>PS00004|PDOC00004|CAMP_PHOSPHO_SITE CAMP- and cGMP-dependent protein kinase phosphorylation site.

Query: 403 RRYT 406
 Query: 988 KRPT 991
 Query: 1630 KKVS 1633

>PS00005|PDOC00005|PKC_PHOSPHO_SITE Protein kinase C phosphorylation site.

Query: 212 TSK 214
 Query: 244 SGR 246
 Query: 247 SRR 249
 Query: 438 TVK 440
 Query: 447 SKR 449
 Query: 475 TGK 477
 Query: 567 SNR 569
 Query: 579 TQR 581
 Query: 720 SAR 722
 Query: 809 TLR 811
 Query: 822 TVR 824
 Query: 915 STK 917
 Query: 945 SER 947
 Query: 959 SPK 961
 Query: 1036 SQR 1038
 Query: 1057 SIR 1059
 Query: 1060 TAK 1062
 Query: 1232 TEK 1234
 Query: 1236 TRR 1238
 Query: 1416 TQK 1418
 Query: 1455 SDK 1457
 Query: 1473 TEK 1475
 Query: 1545 STR 1547
 Query: 1602 TVK 1604
 Query: 1639 SYR 1641

>PS00006|PDOC00006|CK2_PHOSPHO_SITE Casein kinase II phosphorylation site.

Query: 65 TGEE 68

FIG. 7e

Query: 82	TYPD	85
Query: 108	SRLE	111
Query: 144	SFHE	147
Query: 207	SNQD	210
Query: 213	SKKD	216
Query: 247	SRRE	250
Query: 326	TSQE	329
Query: 343	TETE	346
Query: 411	SGLD	414
Query: 551	SPED	554
Query: 688	SDTD	691
Query: 707	SSVE	710
Query: 713	TSGE	716
Query: 730	SSDD	733
Query: 753	SDSE	756
Query: 809	TLRD	812
Query: 880	SKQD	883
Query: 943	TASE	946
Query: 991	TATE	994
Query: 1020	TNVD	1023
Query: 1180	TIYE	1183
Query: 1236	TRRE	1239
Query: 1395	SSCD	1398
Query: 1435	SQEE	1438
Query: 1455	SDKE	1458
Query: 1570	SEIE	1573
Query: 1639	SYRD	1642

>PS000007|PDOC00007|TYR_PHOSPHO_SITE Tyrosine kinase phosphorylation site.

Query: 246	RSRRERQY	253
Query: 811	RDTIDQGLY	819
Query: 830	REILDGLAY	838
Query: 1444	RHHEITY	1450
Query: 1617	KLVCDEIY	1624

>PS000008|PDOC00008|MYRISTYL N-myristoylation site.

Query: 218	GGHRTA	223
Query: 384	GVSLAA	389
Query: 494	GQECGE	499
Query: 599	GAFGAV	604
Query: 613	GCCYAV	618
Query: 684	GQPASD	689

FIG. 7f

Query: 742 GGVFSQ 747

Query: 782 GCHESE 787

Query: 900 GMVGTA 905

Query: 914 GSTKSA 919

Query: 1373 GVSIAI 1378

Query: 1514 GSFSNA 1519

>PS00009|PDOC00009|AMIDATION Amidation site.

Query: 475 TGKK 478

>PS00107|PDOC00100|PROTEIN_KINASE_ATP Protein kinases ATP-binding region signature.

Query: 596 LGKGAFGAV 604

>PS00108|PDOC00100|PROTEIN_KINASE_ST Serine/Threonine protein kinases active-site signature.

Query: 844 MIHRDLKPVNIFL 856

>PS00116|PDOC00107|DNA_POLYMERASE_B DNA polymerase family B signature.

Query: 687 ASDTDGLDS 695

Protein Family / Domain Matches, HMMer version 2

Searching for complete domains in PFAM
hmmfam - search a single seq against HMM database
HMMER 2.1.1 (Dec 1998)
Copyright (C) 1992-1998 Washington University School of Medicine
HMMER is freely distributed under the GNU General Public License (GPL).

HMM file: /prod/ddm/seqanal/PFAM/pfam4.4/Pfam
Sequence file: /prod/ddm/wspace/orfanal/oa-script.16895.seq

Query: 14790

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
pkina	Eukaryotic protein kinase domain	282.0	7.8e-81	4
Ribosomal_L23	Ribosomal protein L23	5.0	3.9	1
mRNA_cap_enzyme	mRNA capping enzyme	-181.3	9.6	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
pkina	1/4	332	443 ..	30	134 ..	69.4	3.7e-18
pkina	2/4	501	539 ..	237	278 .]	22.8	3.2e-05
pkina	3/4	590	662 ..	1	66 [.	51.2	4.2e-13
pkina	4/4	797	1001 ..	65	278 .]	138.7	1.1e-37
Ribosomal_L23	1/1	1223	1244 ..	1	23 [.	5.0	3.9
mRNA_cap_enzyme	1/1	1004	1309 ..	1	504 [.]	-181.3	9.6

Alignments of top-scoring domains:

pkina: domain 1 of 4, from 332 to 443: score 69.4, E = 3.7e-18

*->ilk..kesls..lrEigilkrIsHpNivrllgvfed...tddhlylv
+++ K+++++ +E 1 +lsHpN+vr+l + +++++ + +

14790 332 KIDKcKKQIQgtETEFNSLVKLSHPNVVRYLAMNLKeqdDSIVVDIL 378

mEymegGdLfdylrrngplsekeakkialQilrGleYLHsngivHRDLKp
E+ g +L+ +l+ gp++ ++++++ Q+l+Gl+YLHsn++vH L +

14790 379 VEHISGVSLAAHLSHSGPIPVHQLRRYTAQLLSGLDYLHSNSVVKVLSA 428

eNILldengtvKiaD<-*
+N+L+d +gtvKi+d

14790 429 SNNVLVDAEGTVKtTD 443

http //tango mpi com/seqanal/orfanal/14790 16890 html

FIG. 7g

pkina: domain 2 of 4, from 501 to 539: score 22.8, E = 3.2e-05
 ->rlplpsncSeelkdLlkkcLnkDPskRpGsatakeilnhpwf<-
 + +ps ++ +++d+lkkc ++D ++R+ + ++l+h ++
 14790 501 PVTIPSDLPAQDFLKKCVCLDDKERW---SPQQLKHSFI 539

pkina: domain 3 of 4, from 590 to 662: score 51.2, E = 4.2e-13
 *->yelleklGeGsfGkVykakhk.tgkivAvKilkkels.....lrE
 +e l+ lG+G+fG V k+++k +g+ +AvK ++ s++ ++ E
 14790 590 FEELQLLGKGAFAVIKVNKLGGCCYAVKRIPINPASrqfrriKGE 636
 iqilkrIsHpNivrlIlgvfedtdhl<-*
 + +l rl+H+Niv++ ++ +++++
 14790 637 VTLLSRLHHENIVRYNNAWIERHERP 662

pkina: domain 4 of 4, from 797 to 1001: score 138.7, E = 1.1e-37
 *->hlylvMeymegGdLfdyIrrngplsekeakkialQilrGleYlHsng
 +ly+ mEy+e+ L+d + + + + + +++++il Gl+Y+H++g
 14790 797 YLYIQMEYCEKSTLRDTIDQGLYRDTVRLWRLFREILDGLAYIHEKG 843
 ivHRDLKpeNIlldengtvKiaDFGLArll.....
 +HRDLKp NI+ld++ +vKi+DFGLA++ + ++++++++ +
 14790 844 MIHRDLKpVNIFLDSDDHVKIGDFGLATDHLafsadskqddgtgdlksd 893
 ..eklttfvGTpwyMmAPEvileg...rgysskvDvWSlGviLyElltgg
 ++++lt+ vGT Y+ +PEV +g++++ y kvD SlG+i++E+
 14790 894 psGHLTGMVGTALYV-SPEV--QGstkSAYNQKVDLFLSLGIIFFEMS--- 937
 plfpgadlpafgtggdevdqliifvklPfsdelpktridpleelfrikkr
 P e++f ++
 14790 938 -----YHPMV-----TASERIFVLNQL 954
rlplpsncSee....lkdLlkkcLnkDPskRpGsatakeilnhpwf
 ++++ ++p+ + + + + k+++++Ln DP+kRp ta+e+l++ +
 14790 955 rdpSPKFPEDFDGehakQKSVISWLLNHDPAKRP---TATELLKSELL 1001
 <-*
 14790 - -

Ribosomal_L23: domain 1 of 1, from 1223 to 1244: score 5.0, E = 3.9
 ->tdiikyPviTeKlamnllepnk<-
 +iil y +TeKl++++e ++
 14790 1223 VYIILYDAVTEKLTREVEA-KF 1244

mRNA_cap_enzyme: domain 1 of 1, from 1004 to 1309: score -181.3, E = 9.6
 *->nqtterVvylhkiElfsvpelngKKiglgikLpktddeslrtnVakl
 q++e + e+ l L+++d++ rtm +
 14790 1004 PQMEES-----ELHEV-----LHHTLTNVGKAYRTMMAQI 1034
 lglamtktktfPddeGsqPVsferkdleesLkekdyfveKTDGircshgF
 + P + + + + + vce lr+
 14790 1035 FS-----QRISPAIDYTYDSILKGNFSIRTAKMQHVCEI--IRI---- 1074
 NRTGFLIAaLlFlvehpgleeeaiSHLSgef..lidReknyYKQDYIDlI
 + +h + + + + l+ R +Y ++
 14790 1075 -----FKRHGAVQL-----CTPlLLPRNRQIY-----EHN 1099
 pkrIfPrekdktkakeptyhrgtllDGEIvidinriaveqkTlrYvVFD
 ++ lf + + l++ l d ++ rYv +
 14790 1100 EAALFMD-----HSGMLVM-LPFD----LRIPF-ARYVARN 1129
 alaisGqtvIqrd.lskrLgdefikavkKpfdefkvvmpdakilnqqkYN
 + ++ + lr + L+ f +e+ d++++
 14790 1130 NILNLKRYCIERVFPRKLDLDR-F-----HPKELLECAFDIVTSTT--- 1168
 FpfkiglkHmslsyggqlklkaeskmviskadampkllHindGLiftcvr
 f++ + ++y ++ + a +
 14790 1169 NSFLPTAEIITYIYIEIQEPPA-----L-----QE- 1193
 dtpyieGeilVEPGNSYldfnLLKWKpkeenTvDFelilefeevndPeld
 ++++i+ + Llk +l + +
 14790 1194 RNYSIYL-----NHTMLLKA-----ILLHCG-----IP 1216
 ekdgfslyLdYdampGELfkfslgVWqgGfnkrFevihdtdqiffrvafqk
 e++ + +++++ + +++++ v++
 14790 1217 EDK-----LSQVYIILY-DAVTEKLTRE-----VEAK- 1243
 lgRlkhefaelsVsdkdwyklkaleqpIdGrIVEclradieilIFQegrW
 f +ls+s +l + +E ++ +
 14790 1244 -----FCNLSLSSNSLCRLY-----KFIEQKGLDLDQ----- 1269
 eylrfRdDKqgalktgGYsgNhstvekvllsikDgvsieElklfpGmY
 + ++ + q+ tg + + + + + +E+ l

<http://tango.mpi.com/seqanal/orfanal/14790.16890.html>

FIG. 7h

14790 1270 LMPTINSLIKQK--TG-----IAQLVKYGLKD---LEEVVGLLK--- 1303

```
FAGAktlikr<--*
```

 $k + ik$

14790 1304 ----KLGIKL 1309

//

Searching for complete domains in SMART

hmmpfam - search a single seq against HMM database

HMMER 2.1.1 (Dec 1998)

Copyright (C) 1992-1998 Washington University School of Medicine

HMMER is freely distributed under the GNU General Public License (GPL).

```
HMM file: /ddm/robison/smart/smart/smart.all.hmms
```

```
Sequence file: /prod/ddm/wspace/orfana1/oa-script.16895.seq
```

Query: 14790

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
serkin_6		184.2	2.2e-51	2
tyrkin_6		-40.3	2e-09	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
serkin_6	1/2	286	539 ..	1	231 []	17.2	2.2e-11
serkin_6	2/2	590	1001 ..	1	231 []	166.8	3.5e-46
tyrkin_6	1/1	590	1001 ..	1	280 []	-40.3	2e-09

Alignments of top-scoring domains:

serkin_6: domain 1 of 2, from 286 to 539: score 17.2, E = 2.2e-11

```

14790 286      *-->YellkkIgKgaFGkVylardkktgrlvAiKvik.....
      ++Igk      Vy a ++ tg v  + + + +++ ++
      IGSDEQLGK----LVYNALETtgGTFV---LLYewlqwqkkmgpfl 325

      .....erilrEikiLkk.dHPNIVkLydvfed.....dkl
      ++++++ ++ +++ + ++ E + Lk HPN+V+++ + +++++ +
14790 326      tsqekekidckckkqiQGTETEFNSLVK1SHPNVVRYLAMNLkeqgdsIVV 375

      ylVmEyceGdlGdLfdlkkrgrrgrlkrvlsE.earfyfrQilsaLeYlH
      + E+++G +L +l +g+ ++ ++ r+y++Q+1s+L+YLH
14790 376      DILVEHISG---VSLAAHLSHSGP-----IPVhQLRRYTAQLLSGLDYLH 417

      sqgIiHRDLKPeNiLLds..hvkLaDFGLArql.....ttfvgTp
      S+ ++H L +N+L+d +++vK+D +++++l + +++ + t v
14790 418      SNSVVHKVLSASNVLVDAegTVK1ITDYSISKRLadickedvfeQTRV--- 464

      eYmAPeVl.....gYgkpvad:WSLGCilyEl1tGkpPFPqldl1fkk
      +++++ + +gk +D+WG 1G +l 1 G+ + +
14790 465      -----RFsdnalpyKTGK-KGDVWRGLLLLLSLSQGQE-----CGEYPV 502

      ig....SpeakdL1kk1LvkdPekRlta.eaLedeldikaHPff<-*
      +++ + ++ d+kkk+ + d +R++ ++L+ H +f
14790 503      Ttpsd1PAFDPLKKVCVLDKDERWSPqQLLK-----HSFI 539

```

```
serkin_6: domain 2 of 2, from 590 to 1001: score 166.8, E = 3.5e-46
```

```

      *->YellkkIGkGAFgKvYlarkdkktgrlvAikvik.....
          +e l+l+IGkGAFg V ++++k +g +A+K+i+ ++ +++ ++ +++
14790    590   FEELQLLGKGAFGAIVIKVQNKLGGCCYAVKRIPanpasrqfrakge 636

      .....
          + ++ ++++ + + ++++++ ++++++++ +++++ +++++
14790    637 vtllsrllhenivryynawierherpagpgtpppsdgplakddraargap 686

      .....
          +++++ ++ + +++ ++ ++++++ + + ++++++++
14790    687 asdtdgldsveaaappilsssvewstsgersasarfpatgggsddedd 736

      .....erilrEikiLkk...dHPNIvKLydv
          +++++++ ++ ++++++ + E ++++++d+ +++++
14790    737 dedehggvfsgsfldpsdsediiFDNEDENSKSQNGdeDCNEKNKGCHES 786

fed.....dklylvmEyceGdlGdLfdllkkrgrrglrkvlsE.earfy
          + +++ +ly+ mEyce +L+d +++ + + + +++
14790    787 EPSvtteavHYLYIQMEYCEK--STLRDTIDQGGLY-----RDTvRLWRL 828

frQilsaLeYLHsqgIiHRDLKPENiLLds..hvKlaDFGLArql....
          fr+il+L+Y+H++g iHRDLKP Ni+Lds++hvK++DFGLA+ +
14790    829 FREILDGLAYIEHKGIHRDLKPNIFLDsdhdHVKIGDFGLATDHlafsa 878

      .....ttfvGTpeYmAPEvl....gYgpavDiw
          ++++++++ ++++++ t +vGT Y++PEV +++++ Y+ +vD

```

http://tango.mpi.com/seganal/orfanel/14790_16890.html

FIG. 7i

```

14790 879 dskqddqtdgdlksdpsghlTGMVGTALYVSPEVQgstksAYNQ-KVDLF 927
      SlGcilyElltGkpPFp..qldlifkkig.....Speakd
      SlG+i++E+ + p ++ ++++++++ + + + + + K+
14790 928 SLGIIFFEMSY-HPMVTASERIFVLNQLRdptspkfpddgeHAKQKS 976
      LikklLvkdPekRlta.eaLedeldikaHPff<-*
      i+ lL+ dP+kR+ta+e+L+ + +
14790 977 VISWLLNHDPakRPTaTELLK-----SELL 1001

```

tyrkin_6: domain 1 of 1, from 590 to 1001: score -40.3, E = 2e-09

```

*->ltlgkklGeGaFGeVykgTlk...ieVAVKtLkeda...keeFlrE
      + ++ LG GaFG V k + k ++ AVK + + + + + E
14790 590 FEELQLLGKGAFGAVIKVQNKldgCCYAVKRIPINPasrqFRRIKGE 636
      akiMkklGgkHpNiVklLlGvcteeegrrFmevePlmivmEymegGdLldyL
      ++++l +H+NiV+ + + +e
14790 637 VTLLSRL--HHENIVRYYNawIE----- 657
      rknrpK.....
      r rp ++++++++ + + + + + + + + + + + + + + + + +
14790 658 RHERPAgpgtpppdsgplakddraargqpasdtgdldsveaappilss 707
      .....
      + + + + + + + + + + + + + + + + + + + + + + + + + + + + +
14790 708 sviewstgersasarfpatgpgssddedddedehggvifsqslpasdses 757
      .....
      + + + + + + + + + + + + + + + + + + + + + + + + + + + + +
14790 758 diifdnedensksqngdedcnekngcheseptteavhylyiqmeycek 807
      .....lsldLlslfAlQIAKGMeYLesknfvHRDLAARNcLvG
      ++ + + + + + + L I +G +Y+++k+ +HRDL N+ ++
14790 808 stlrdtidqglyRDTVRLWRLFREILDGLAYIHEKGMiHRDLKPVNIFLD 857
      enkvvKIsDFGLsRdlyddDkkG....eskdYrkkggkggkttlPir.
      ++ +vKI+DFGL+ d+ ++ + + + + +d + + + + + + + + +
14790 858 SDDHVKIGDFGLATDHLAF-SADskqddQTGDLIKSDPSGHLTGMVGTAl 906
      WmAPesl..kdgkFtskSDVWSFGVLWEiftlGegPYpgeiqqfmsnee
      +PE+ ++ + + + + k D +S G+ +E+ Y + + + + +e
14790 907 YVSPEVQgsTKSAYNQKVDLFSLGIIFEM-----SYHPM----VTASE 946
      ...vleylkkGyRlpkPendlpisS.vtCPdelYdlMlqCwaedPedRpt
      + vl++l++ ++ Pe+ ++ + + + + + + + + +dP++Rpt
14790 947 rifVLNQLRDPTSPKFPED----FdDGEHAKQKSVISWLLNHDPakRPT 991
      Fsel..verl<-*
      +el + e+l
14790 992 ATEllKSELL 1001

```

ProDom Matches

FIG. 7j

ProdomId	Start	End	Description	Score
View Prodom 40346 <input type="button" value="Boxer"/> <input type="button" value="Showing match"/> <input type="button" value="Go!"/>	18	323	p99.2 (2) O01712(1) O61651(1) // INITIATION FACTOR KINASE EUKARYOTIC EIF-2 ALPHA EIF-2ALPHA	295
View Prodom 137719 <input type="button" value="Boxer"/> <input type="button" value="Showing match"/> <input type="button" value="Go!"/>	19	1504	p99.2 (1) O74297_NEUCR // CPC3 PROTEIN	87
View Prodom 1 <input type="button" value="Boxer"/> <input type="button" value="Showing match"/> <input type="button" value="Go!"/>	341	1017	p99.2 (2773) CC2(14) KKIT(14) KPC1(13) // KINASE PROTEIN TRANSFERASE ATP-BINDING SERINE/THREONINE-PROTEIN PHOSPHORYLATION RECEPTOR TYROSINE-PROTEIN PRECURSOR TRANSMEMBRANE	107
View Prodom 150228 <input type="button" value="Boxer"/> <input type="button" value="Showing match"/> <input type="button" value="Go!"/>	1105	1206	p99.2 (1) O74297_NEUCR // CPC3 PROTEIN	90
View Prodom 2305 <input type="button" value="Boxer"/> <input type="button" value="Showing match"/> <input type="button" value="Go!"/>	1106	1230	p99.2 (23) SYH(12) // SYNTHETASE AMINOACYL-TRNA HISTIDYL-TRNA PROTEIN LIGASE ATP-BINDING BIOSYNTHESIS HISTIDINE--TRNA HISRS KINASE	137
View Prodom 42726 <input type="button" value="Boxer"/> <input type="button" value="Showing match"/> <input type="button" value="Go!"/>	1274	1648	p99.2 (2) O01712(1) O61651(1) // INITIATION FACTOR KINASE EUKARYOTIC EIF-2 ALPHA EIF-2ALPHA	221
ProdomId	Start	End	Description	Score

View Prodom 40346

40346 p99.2 (2) O01712(1) O61651(1) // INITIATION FACTOR KINASE EUKARYOTIC
EIF-2 ALPHA EIF-2ALPHA
Length = 296

Score = 295 (108.9 bits), Expect = 1.7e-28, Sum P(2) = 1.7e-28
Identities = 58/144 (40%), Positives = 95/144 (65%)

Query: 18 SYPQRQDHELQALEAIYGADFDLRPDACGPVKPEPEINLVLYP--QGLTGEEVYVKVDL 75
S+ +RQ EL+ +++I+G D +DLRP A + +P +I + L P G E YV L
Sbjct: 9 SFRERQAELEVIKSIIFGCDVEDLRPQANPSLWKPTDIRIQLTPLRDSSNGLETIVCTKL 68

Query: 76 RVKCPPTYPDVVPPEIELKNAKGLSNESVNLKSRLEELAKKHCGEVMIFELAYHVQSFLS 135
V CP YP + P+I L+ +KG+S++ + L+++L+ +++ GEVMI+ELA VQ+FL
Sbjct: 69 HVTCPSPYKLPKPKISLEESKGMSDQLLEALRNQLQAQSQELRGEVMIELAQTVAFL 128

Query: 136 EHNKPPPKSFHEMLERRAQEEQQ 159
EHNKPP SF+++ML+ + + +Q+
Sbjct: 129 EHNKPPKGSFYDQMLQDKQRDQE 152

Score = 63 (27.2 bits), Expect = 1.7e-28, Sum P(2) = 1.7e-28
Identities = 15/59 (25%), Positives = 30/59 (50%)

Query: 266 EILYFN-MGSPDQLMVHKGKIGSDEQLGKLVYNALETATGGFVLLYEWVLOWQKKMGP 323
E LYF+ MG + +G C+G ++ G + Y ++ G + + EW +++ + P
Sbjct: 230 ETLYFHKMGR---QIQRGCCVGHSSQR-GCIAYTGIDMHCGQLLYITENIKYSQLEQP 283

View Prodom 42726

>42726 p99.2 (2) O01712(1) O61651(1) // INITIATION FACTOR KINASE EUKARYOTIC
EIF-2 ALPHA EIF-2ALPHA
Length = 469

Score = 221 (82.9 bits), Expect = 7.5e-15, P = 7.5e-15
Identities = 92/341 (26%), Positives = 160/341 (46%)

FIG. 7k

Query: 1274 INSLIKQKTGIAQLVKYGLKDLKLEXXXXXXXXXXXXXXXXXNGLVKVQQ--HNGIIFQF 1331
+ SL++ K A L + L++LE GL + + GI++Q
Sbjct: 70 LKSLMRGKGGAASLARGALRELETVVGLAYSLGKCPHIHWAGLPISFDRASNGGIVWQM 129

Query: 1332 VAFIKRRQRAVPEILAAGGRYDLIPQF-RGPQALGPVPTAIGV-----SIAIDKISAA 1384
A +K + P +LA G RYD ++ +F + Q P A GV + ++DK+ AA
Sbjct: 130 TADLKNRSGHPSVLAIGERYDMLHEFQKQAQKFNPPAMPARGVLSGAGLTFSLDKLVAA 189

Query: 1385 VLNMEESVTISSCDLLVSVGQMSMSRAINLTQKL-WTAGITAEIMYDWSQSQEELQEYC 1443
V +E + + D+ + G + + +L W+ GI I+ S+ +E Q+
Sbjct: 190 V-GVEYAKDCRAIDVGICVCGTRPPLKDVITYIMRLLWSVGIRCGIVEAASELGDEAQDLA 248

Query: 1444 RHHEITYVALVSDKEGSHVVKVSFEKERQTEKRVLETETELVDHVLQKLRTKVTDERNGRE 1503
R + +V LV++ GS ++V+SFE+ER E+ + TELV+ + + LR+ D NG
Sbjct: 249 RLGAH-HVILVAEN-GS-LRVRSFERERFQERHLTRTELVEFIQKMLRS---DGLNGTTV 302

Query: 1504 SDNLAVQNLKGSFNSAS-----GLF-EIHGATV-----VPIVSV--LAPEKLS 1543
DN + + GS N S GL AT+ +P + V L +K +
Sbjct: 303 -DNFSLHLSALGSGDNRSSGGKERERGENGLSTSASNATIKNNYSQPLNLTQVTLTHDKPT 361

Query: 1544 ASTRRRYETQVQTRLQTSANLHQSSEIEILAVDLPKETI 1584
A+ +RR E QV ++ ++L+ +K + + +L V+LP +
Sbjct: 362 ANYKRLENQVAQMSSTLSQFLKKETPV-VLVVELPPAVV 401

Score = 150 (57.9 bits), Expect = 4.2e-07, P = 4.2e-07
Identities = 66/300 (22%), Positives = 137/300 (45%)

Query: 1368 VPTAIGVSTIAIDKISAAVLNMEESVTISSCDLLVSVGQMSMSRAINLTQKL-WTAGITA 1426
V + G++ ++DK+ AAV +E + + D+ + G + + +L W+ GI
Sbjct: 173 VLSGAGLTFSLDKLVAAV-GVEYAKDCRAIDVGICVCGTRPPLKDVITYIMRLLWSVGIRC 231

Query: 1427 EIMYDWSQSQEELQEYCRHHEITYVALVSDKEGSHVVKVSFEK--ERQTEKRVLETETELVD 1484
I+ S+ +E Q+ R + +V LV++ V+ E+ ER + L E +
Sbjct: 232 GIVEAASELGDEAQDLARLGAH-HVILVAENGLRVRSFERERFQERHLTRTEL-VEFIQ 289

Query: 1485 HVLQK--LRTKVTDERNGRE---SDNLAV-----QNLKGSFNSASGLFEIHGATVVI 1533
+L+ L D + A DN + + G ++AS + + +P
Sbjct: 290 KMLRSDGLNGTTVDNFSHLSALGSGDNRSSGGKERERGENGLSTSASNATIKNNYSQPLN 349

Query: 1534 VSV--LAPEKLSASTRRRYETQVQTRLQTSANLHQSSEIEILAVDLPKETILQFLSL- 1590
+ V L +K +A+ +RR E QV ++ ++L+ +K + + +L V+LP + +
Sbjct: 350 LQVTLTHDKPTANYKRLENQVAQMSSTLSQFLKKETPV-VLVVELPPAVVNAIVGAI 408

Query: 1591 -EWDADAEQAFNTTVKQLLSRLPK-QRYLKLVCDEIYNIKVEKKVSVFLYLYSYRDDYYRIL 1648
+ ++ + ++ R K +RY+ + +E+ + + K ++ LYS D YR++
Sbjct: 409 NPREIRKRETEPEINYYVIERFSKYKRYISEINEEVVDYLSDAKTPIVALYSISDSYYRVI 468

View Prodom 2305

Boxer

Showing match

Go

>2305 p99.2 (23) SYH(12) // SYNTHETASE AMINOACYL-TRNA HISTIDYL-TRNA PROTEIN
LIGASE ATP-BINDING BIOSYNTHESIS HISTIDINE--TRNA HISRS KINASE
Length = 145

Score = 137 (53.3 bits), Expect = 3.9e-08, P = 3.9e-08
Identities = 41/131 (31%), Positives = 62/131 (47%)

Query: 1106 DHSGMLVMLPFDLRIPFARYVARNNI---LNLKRYCIERVFRPRK--LDRFHPKELLE 1159
D G L+ L +DL +PFARYVA N + L LKRY I +V+R + + R +E +C
Sbjct: 18 DQGGELLSRLYDLTVPFARYVAMNLLKVNTNPLKRYHIAKVYRRDRPAMTRGRYREFYQC 77

Query: 1160 AFDIVTSTTNSFLPXXXXXXXXXXQEFPAQERNYSIYLNHTMLLKAILLHCGIPEDK 1219
FDI+ ++ P + + N+ I +NH +L ++L P K
Sbjct: 78 DFDII-GEYDTMAPDAEILKILTEILSQLGIRELGNFKIKINHRGILDSLLQ----PWPK 132

Query: 1220 LSQVYIILYDA 1230
Q Y+ Y A
Sbjct: 133 TLQEYLTQYKA 143

View Prodom 1

>1 p99.2 (2773) CC2(14) KKIT(14) KPC1(13) // KINASE PROTEIN TRANSFERASE
ATP-BINDING SERINE/THREONINE-PROTEIN PHOSPHORYLATION RECEPTOR
TYROSINE-PROTEIN PRECURSOR TRANSMEMBRANE
Length = 431

<http://tango.mpi.com/seqanal/orlanal/14790.16890.html>

FIG. 71

Score = 107 (42.7 bits), Expect = 0.0032, Sum P(2) = 0.0032
Identities = 36/106 (33%), Positives = 50/106 (47%)

Query: 848 DLKPVNIFLDSDDH-----VK-IGDFGLATDHLAFSADSKQDDQTGDLIKSDPSGHLTG 900
DLKP NI LD + H +K I DFGLA + +S+ S ++ + +
Sbjct: 220 DLKPENILLDEESHENTPNMIKLIADPGLAKE--IYSSSSTYEEMSSSQAVFGSHQTTST 277

Query: 901 MVGTALYVS-----PEVQGSTKSA-----YNQKVDLFSLGIIFFEM 936
M GT YVS PE SA Y+ K D++S G+I +EM
Sbjct: 278 MCGTPYYVSMKSMAPYMAPESSATNYQKYSTKSDVWSFGVILYEM 323

Score = 105 (42.0 bits), Expect = 3.5e-05, Sum P(3) = 3.5e-05
Identities = 35/102 (34%), Positives = 48/102 (47%)

Query: 824 RLWRLFREILDGLAYIHEK-----GMIHR-----DLKPVNIFLDSDDH-----VK-IGDF 867
+L +I GL Y+H K G+IHR DLKP NI LD + H +K I DF
Sbjct: 187 QLMHYVHQIAKGLLEYLHKNQKHQGIHRAKKVDLKPENILLDEESHENTPNMIKLIADF 246

Query: 868 GLATDHLAFSADSKQDDQTGDLIKSDPSGHLTGMVGTALYVS 909
GLA + +S+ S ++ + + M GT YVS
Sbjct: 247 GLAKE--IYSSSSTYEEMSSSQAVFGSHQTTSTMCGTPYYVS 286

Score = 74 (31.1 bits), Expect = 0.91, Sum P(2) = 0.60
Identities = 30/127 (23%), Positives = 60/127 (47%)

Query: 341 QGTETEFNSLVKLSHPNVVRYLAMNLKEQDDSIIVDIL-VEHISGVSLAAHLSHSGPIPV 399
+G+ E+ + + ++ L++ + +++ + + H+S S + LSHS
Sbjct: 134 EGSLEVMEYMSGGSEDMYMKLSLETVMKIAMMILQFMQIMHMSSESES--LSHS----- 186

Query: 400 HQLRRYTAQLLSGLDYLHSNS-----VVHKV---LSASNVLVDAEG-----TVK-ITD 443
QL Y Q+ GL+YLS + ++H+ L N+L+D E +K I D
Sbjct: 187 -QLMHYVHQIAKGLLEYLHKNQKHQGIHRAKKVDLKPENILLDEESHENTPNMIKLIAD 245

Query: 444 YSISKRL 450
+ ++K +
Sbjct: 246 FGLAKEI 252

Score = 65 (27.9 bits), Expect = 3.5e-05, Sum P(3) = 3.5e-05
Identities = 29/124 (23%), Positives = 50/124 (40%)

Query: 907 YVSPEVQGSTKSAYNQKVDLFSLGIIFFEM-SYHPMTASERIFVLNQLRDPSPKFPED 965
Y++PE + Y+ K D++S G+I +EM + P E +++ S K E
Sbjct: 294 YMAPESSATNYQKYSTKSDVWSFGVILYEMLTGKPPFPFGES--EVSEEEFYQSMKNMEV 351

Query: 966 FDDGEHAKQKSVISWLLNHDPAKRPT-----ATELLKSELLPPPMEESELHE 1013
+ G + V+S ++ + P A +LLK L P+ E
Sbjct: 352 LEMGPREETIQKVMKIVEKKGERMPQPSSSNCEVVSQEAKDLLKKCLQKDPEKRRPTFEE 411

Query: 1014 VLHH 1017
+L H
Sbjct: 412 ILQH 415

Score = 55 (24.4 bits), Expect = 3.5e-05, Sum P(3) = 3.5e-05
Identities = 12/23 (52%), Positives = 18/23 (78%)

Query: 589 EFEELQ-LLGKGAFGAVIKVQNK 610
++E L+ LLGKG+FG V K ++K
Sbjct: 33 QYELLKKLLGKGSFGKVYAKHK 55

Score = 49 (22.3 bits), Expect = 34., Sum P(2) = 1.0
Identities = 12/39 (30%), Positives = 24/39 (61%)

Query: 507 DLPADFDQFLKKCVCLD-DKERWSPQQLLKHS-FINPQP 543
++ + +D LKKC+ D +K R + +++L+H F+ P
Sbjct: 385 EVSQEAKDLLKKCLQKDPEKRRPTFEEILQHPWFLMRNP 423

Score = 40 (19.1 bits), Expect = 0.0010, Sum P(3) = 0.0010
Identities = 8/11 (72%), Positives = 9/11 (81%)

Query: 596 LGKGAFGAVIK 606
LG G+FGAV K
Sbjct: 2 LGTGSFGAVYK 12

View Prodom 150228

>150228 p99.2 (1) 074297_NEUCR // CPC3 PROTEIN
Length = 108

FIG. 7m

<http://tango.mpi.com/seqanal/orfanal/14790.16890.html>

Score = 90 (36.7 bits), Expect = 0.0039, P = 0.0039
Identities = 32/105 (30%), Positives = 43/105 (40%)

Query: 1105 MDHSGMLVMLPFDLRIPPARYVAR--NNILNLKRYCIERVFRPRKLDLRFHPKELLECAFD 1162
+D +G ++ LPFDL + AR +AR N+ + K Y +FR R P E FD
Sbjct: 3 LDQNGTVLQLPFDLMMGHARSLARITNSPVVQKSYSGNIFRDRH-GGGQPDVYGEVD 61

Query: 1163 IVTSTTNSF-LPXXXXXXXXXXXXQEFFPALQERNYSIYLNHTMLL 1206
IVTS L FP + L H+ LL
Sbjct: 62 IVTSDALDLALKEAEVIKVLDEIATAFPTVSSTPICFQLGHSDDL 106

View Prodom 137719

Boxer

Showing match

Go

>137719 p99.2 (1) 074297_NEUCR // CPC3 PROTEIN
Length = 304

Score = 87 (35.7 bits), Expect = 0.61, Sum P(2) = 0.45
Identities = 41/156 (26%), Positives = 64/156 (41%)

Query: 19 YPQRQDHELQALEIYGADFQDLRPDAGPVPKEPPEINLVLPQGLTGEEVYVKVDLRVK 78
Y + Q+ E+ L+AIYG DF K P ++ + P + +E+ V L V
Sbjct: 44 YQEVQSEVMVLQAIYGEDFTQHEAAHGAWQKSEPRFDIKIKPS--SDQEL--SVTLGVV 99

Query: 79 CPPTYDPDVPEIELKNAKGLSNESVNLLKSRLEELAKK---HCGEVMIFELAYHVQSFLS 135
TYP P + +K+ L ES + E K + MI ++ ++ L
Sbjct: 100 MVATYPKTPPLLTIKDDHSL-RESTKPKIQKFVETQPKIYAQAEQEMIDQIVEGIRDILE 158

Query: 136 E--HNK-----PP---PKSPHEEMLERRAQEEQQR 160
E K P ++ HE L R AQ E++R
Sbjct: 159 EAAQKKVQGLEIPSLEEEERAAHEALARLAQSEKER 194

Score = 49 (22.3 bits), Expect = 0.61, Sum P(2) = 0.45
Identities = 13/48 (27%), Positives = 27/48 (56%)

Query: 1458 EGS HVKVSFEKERQTEKRVLETELVDHVLQK-LRTKVTDERNGREAS 1504
E ++ EKER+ K++ E++ + VL+ L+ ++ +RN + S
Sbjct: 181 EAELARLAQSEKEREERKKLEESKEEERVLQEDMLQEELKRQRNKAKES 228

Score = 47 (21.6 bits), Expect = 0.96, Sum P(2) = 0.62
Identities = 15/61 (24%), Positives = 26/61 (42%)

Query: 238 KHRANSSGRSRERQYSVCNSEDSPGSC-EILYFNMGSPDQLMVHKGCIGSDEQLGKLV 296
+++A S + R Q S + PG E L F+ P ++ G + +GK V
Sbjct: 222 RNKAKESRKKNRSHQLSPDRAPQDPGETDETLMFDQ--PCKITDGSIGNALFFQTVIGKTV 279

Query: 297 Y 297
+
Sbjct: 280 F 280

Score = 47 (21.6 bits), Expect = 0.96, Sum P(2) = 0.62
Identities = 19/83 (22%), Positives = 33/83 (39%)

Query: 750 LPASDSESIIFDNEDENSKSQ-NQDEDCNEKNGCHESEPSVTTEAVHYLIQMEYCEKS 808
L S E ++ D E K Q N+ ++ +KN H+ P + + + +
Sbjct: 200 LEESKEEERVLQEDMLQEELKRQRNKAKESRKKNRSHQLSPDRAPQDPGETDETLMFDQPC 259

Query: 809 TLRDTIDQGLYRDTVRLWRLFRE 831
+ D L+ TV +FRE
Sbjct: 260 KITDGSIGNALFFQTVIGKTVFRE 282

FIG. 7n

14790 Expression in HBV+ Livers

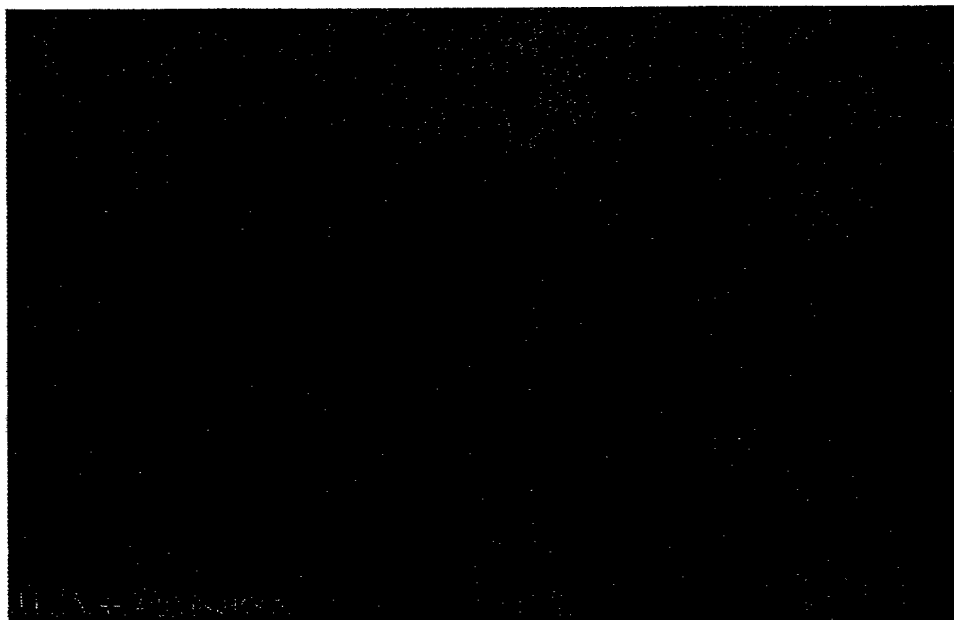


FIG. 8a



FIG. 8b

14790 Expression in Hepatocytes of HCV+ Livers

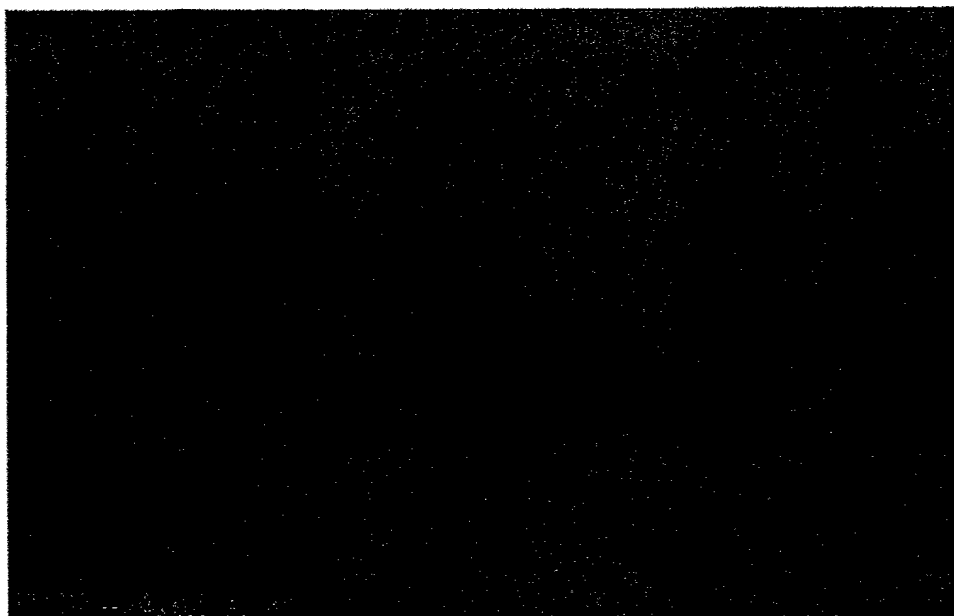


FIG. 9a

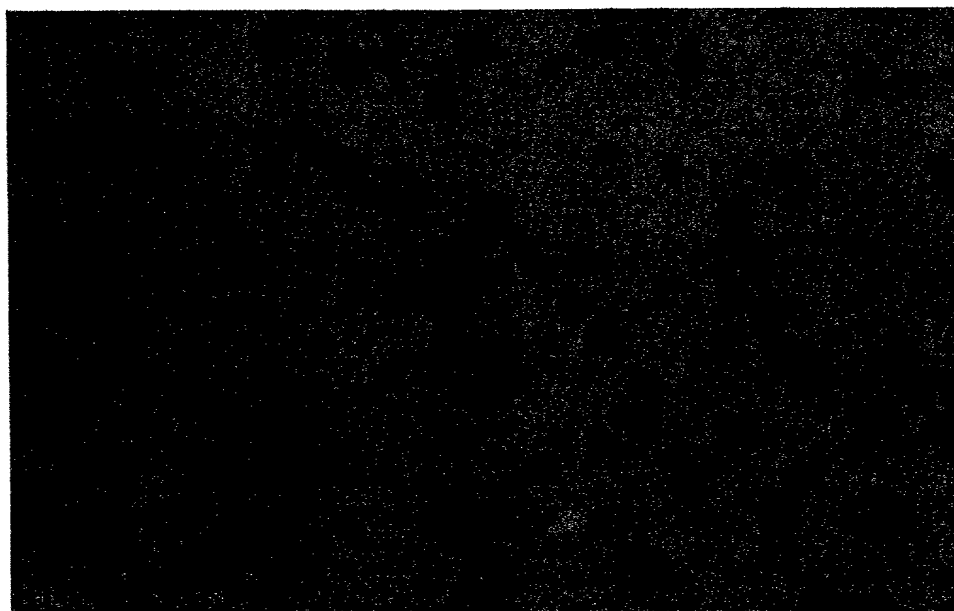


FIG. 9b

DECLARATION FOR UTILITY PATENT APPLICATION

AS A BELOW-NAMED INVENTOR, I HEREBY DECLARE THAT:

My residence, post office address, and citizenship are as stated below next to my name.

I believe I am the original, first and sole inventor of the subject matter which is claimed and for which a patent is sought on the invention entitled: 14790, A NOVEL PROTEIN KINASE MOLECULE AND USES THEREFOR, the specification of which is attached hereto unless the following box is checked:

☒ was filed on February 29, 2000 as United States Application Serial No. _____ and was amended on _____ applicable).

I HEREBY STATE THAT I HAVE REVIEWED AND UNDERSTAND THE CONTENTS OF THE ABOVE-IDENTIFIED SPECIFICATION, INCLUDING THE CLAIMS, AS AMENDED BY ANY AMENDMENT REFERRED TO ABOVE.

I acknowledge the duty to disclose information which is material to the patentability as defined in 37 C.F.R. § 1.56.

I hereby claim foreign priority benefits under 35 U.S.C. § 119(a)-(d) or § 365(b) of any foreign application(s) for patent or inventor's certificate, or § 365(a) of any PCT International application which designated at least one country other than the United States listed below and have also identified below, by checking the box, any foreign application for patent or inventor's certificate, or PCT International application having a filing date before that of the application on which priority is claimed:

Application No.	Country	Date of Filing (day/month/year)	Priority Claimed?
			<input type="checkbox"/> Yes <input type="checkbox"/> No

I hereby claim benefit under 35 U.S.C. § 119(e) of any United States provisional application(s) listed below:

Application Serial No.	Filing Date

I hereby claim the benefit under 35 U.S.C. § 120 of any United States application(s), or § 365(c) of any PCT International application designating the United States, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International application in the manner provided by the first paragraph of 35 U.S.C. § 112, I acknowledge the duty to disclose information which is material to patentability as defined in 37 C.F.R. § 1.56 which became available between the filing date of the prior application and the national or PCT International filing date of this application.

Application Serial No.	Filing Date	Status
		<input type="checkbox"/> Patented <input type="checkbox"/> Pending <input type="checkbox"/> Abandoned

I hereby appoint the following attorneys and agents to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith:

Sanjay S. Bagade (Reg No. 42,280)	Erwin J. Basinski (Reg No. 34,773)
Richard R. Batt (Reg No. 43,485)	Frank P. Becking (Reg No. 42,309)
Kimberly A. Bolin (Reg No. 44,546)	Timothy J. Bortree (Reg No. 43,506)
Sean Brennan (Reg No. 39,917)	Barry E. Bretschneider (Reg No. 28,055)
Tyler S. Brown (Reg No. 36,465)	Nicholas Buffinger (Reg No. 39,124)
Jingming Cai (Reg No. P-44,579)	Alan W. Cannon (Reg No. 34,977)
Mark R. Carter (Reg No. 39,131)	Robert K. Cerpa (Reg No. 39,933)
Thomas Chuang (Reg No. P-44,616)	Thomas E. Ciotti (Reg No. 21,013)
Steven X. Cui (Reg No. 44,637)	Matthew M. D'Amore (Reg No. 42,457)
Raj S. Davé (Reg No. 42,465)	Stephen C. Durant (Reg No. 31,506)
Carolyn A. Favorito (Reg No. 39,183)	Hector Gallegos (Reg No. 40,614)
Debra J. Glaister (Reg No. 33,888)	Kenneth R. Glick (Reg No. 28,612)
Franklin Y. Han (Reg No. 41,055)	Douglas Hodder (Reg No. 41,840)
Charles D. Holland (Reg No. 35,196)	Peter Hsieh (Reg No. P-44,780)
Madeline I. Johnston (Reg No. 36,174)	Richard D. Jordan (Reg No. 33,519)
Ararat Kapouytian (Reg No. 40,044)	Phanesh B. Koneru (Reg No. 40,053)
Kawai Lau (Reg No. 44,461)	Elaine K. Lee (Reg No. 41,936)
Susan K. Lehnhardt (Reg No. 33,943)	Richard H. Lilley (Reg No. 42,803)
David C. Liu (Reg No. P-43,755)	Wen Liu (Reg No. 32,822)
David C. Lundmark (Reg No. 42,815)	Harry J. Macey (Reg No. 32,818)
Thomas D. Mays (Reg No. 34,524)	Gladys H. Monroy (Reg No. 32,430)
Philip A. Morin (Reg No. P-45,926)	Kate H. Murashige (Reg No. 29,959)
Catherine M. Polizzi (Reg No. 40,130)	William C. Revelos (Reg No. 42,101)
Paul J. Riley (Reg No. 38,596)	Robert Saltzberg (Reg No. 36,910)
Debra A. Shetka (Reg No. 33,309)	Kevin R. Spivak (Reg No. 43,148)
Suzannah K. Sundby (Reg No. 43,172)	Kenneth K.D. Vu (Reg No. P46,323)
E. Thomas Wheelock (Reg No. 28,825)	Thomas G. Wiseman (Reg No. 35,046)
Frank Wu (Reg No. 41,386)	Karen R. Zachow (Reg No. P-46,332)

and:

Please direct all communications to:

Carolyn A. Favorito
Morrison & Foerster LLP
2000 Pennsylvania Avenue, N.W.
Washington, D.C. 20006-1888

Please direct all telephone calls to Carolyn A. Favorito at (202) 887-1546.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these

Variable	Mean	Standard deviation	Minimum	Maximum
Age	35.2	10.5	20	55
Gender	0.5	0.5	0	1
Marital status	0.6	0.5	0	1
Education	12.5	2.5	9	16
Income	15.2	5.5	10	25
Health status	0.8	0.4	0	1
Smoking status	0.3	0.5	0	1
Alcohol consumption	0.2	0.4	0	1
Exercise frequency	0.5	0.5	0	1
Stress level	0.7	0.5	0	1
Depression score	0.4	0.5	0	1
Life satisfaction	0.6	0.5	0	1
Quality of life	0.7	0.5	0	1
Healthcare utilization	0.5	0.5	0	1
Health insurance status	0.9	0.3	0	1
Healthcare access	0.8	0.4	0	1
Healthcare cost	0.6	0.5	0	1
Healthcare quality	0.7	0.5	0	1
Healthcare satisfaction	0.8	0.4	0	1
Healthcare accessibility	0.9	0.3	0	1
Healthcare affordability	0.7	0.5	0	1
Healthcare effectiveness	0.8	0.4	0	1
Healthcare safety	0.9	0.3	0	1
Healthcare equity	0.6	0.5	0	1
Healthcare transparency	0.7	0.5	0	1
Healthcare accountability	0.8	0.4	0	1
Healthcare integrity	0.9	0.3	0	1
Healthcare honesty	0.7	0.5	0	1
Healthcare trustworthiness	0.8	0.4	0	1
Healthcare reliability	0.9	0.3	0	1
Healthcare consistency	0.7	0.5	0	1
Healthcare predictability	0.8	0.4	0	1
Healthcare stability	0.9	0.3	0	1
Healthcare security	0.7	0.5	0	1
Healthcare privacy	0.8	0.4	0	1
Healthcare confidentiality	0.9	0.3	0	1
Healthcare professionalism	0.7	0.5	0	1
Healthcare competence	0.8	0.4	0	1
Healthcare skillfulness	0.9	0.3	0	1
Healthcare knowledge	0.7	0.5	0	1
Healthcare experience	0.8	0.4	0	1
Healthcare expertise	0.9	0.3	0	1
Healthcare proficiency	0.7	0.5	0	1
Healthcare capability	0.8	0.4	0	1
Healthcare ability	0.9	0.3	0	1
Healthcare capacity	0.7	0.5	0	1
Healthcare potential	0.8	0.4	0	1
Healthcare power	0.9	0.3	0	1
Healthcare strength	0.7	0.5	0	1
Healthcare influence	0.8	0.4	0	1
Healthcare impact	0.9	0.3	0	1
Healthcare effect	0.7	0.5	0	1
Healthcare effectness	0.8	0.4	0	1
Healthcare effectivity	0.9	0.3	0	1
Healthcare effectualness	0.7	0.5	0	1
Healthcare effectiveness	0.8	0.4	0	1
Healthcare effectivity	0.9	0.3	0	1
Healthcare effectualness	0.7	0.5	0	1
Healthcare effectiveness	0.8	0.4	0	1
Healthcare effectivity	0.9	0.3	0	1
Healthcare effectualness	0.7	0.5	0	1
Healthcare effectiveness	0.8	0.4	0	1
Healthcare effectivity	0.9	0.3	0	1
Healthcare effectualness	0.7	0.5	0	1
Healthcare effectiveness	0.8	0.4	0	1
Healthcare effectivity	0.9	0.3	0	1
Healthcare effectualness	0.7	0.5	0	1
Healthcare effectiveness	0.8	0.4	0	1
Healthcare effectivity	0.9	0.3	0	1
Healthcare effectualness	0.7	0.5	0	1
Healthcare effectiveness	0.8	0.4	0	1
Healthcare effectivity	0.9	0.3	0	1
Healthcare effectualness	0.7	0.5	0	1
Healthcare effectiveness	0.8	0.4	0	1
Healthcare effectivity	0.9	0.3	0	1
Healthcare effectualness	0.7	0.5	0	1
Healthcare effectiveness	0.8	0.4	0	1
Healthcare effectivity	0.9	0.3	0	1
Healthcare effectualness	0.7	0.5	0	1
Healthcare effectiveness	0.8	0.4	0	1
Healthcare effectivity	0.9	0.3	0	1
Healthcare effectualness	0.7	0.5	0	1
Healthcare effectiveness	0.8	0.4	0	1
Healthcare effectivity</				

Date:	Name:	Rosana KAPPELLER-LIBERMANN
	Residence:	Chestnut Hill, MA 02467
	Citizenship:	BR
	Post Office Address:	86 Beacon Street, Chestnut Hill, MA 02467